colección **textos**

ÁLGEBRA MATRICIAL CON APLICACIONES EN ESTADÍSTICA

2a. edición

José Alfredo Jiménez M.

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE BOGOTÁ
FACULTAD DE CIENCIAS

José Alfredo Jiménez M.

Matemático con maestría en Ciencias Estadísticas de la Universidad Nacional de Colombia. Magíster en Banca y Finanzas de la Universidad de Valencia, España. Profesor del Departamento de Matemáticas de la Universidad Nacional de Colombia por más de 10 años.

PUBLICACIONES RECIENTES DE LA FACULTAD DE CIENCIAS

Javier Fernando Cardona
Óscar Roberto Blanco
ELECTRÓNICA DIGITAL Y SU APLICACIÓN A LA INSTRUMENTACIÓN
[guía de laboratorio]
Colección notas de clase

Marco Fidel Suárez
ELECTROQUÍMICA FÍSICA E INTERFACIAL
Colección textos

Sergio Yáñez
Nelfi González
José Alberto Vargas
CARTAS DE CONTROL T^2 MULTIVARIADAS USANDO R y SAS
Colección notas de clase

Liliana Blanco
PROBABILIDAD
2da. edición
Colección textos

Camilo Ernesto López (Ed.)
FUNDAMENTOS Y TÉCNICAS BÁSICAS EN BIOLOGÍA MOLECULAR
Colección notas de clase

Liliana López-Klein
BIOESTADÍSTICA
Colección notas de clase

Álgebra matricial con aplicaciones en Estadística

José Alfredo Jiménez Moscoso

Álgebra matricial con aplicaciones en Estadística

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE BOGOTÁ
FACULTAD DE CIENCIAS

Bogotá D. C. Colombia, junio de 2012

© Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Matemáticas
© José Alfredo Jiménez Moscoso

ILUSTRACIÓN PORTADA
Y CONTRAPORTADA
Profesor Gustavo Rubiano
Departamento de Matemáticas

ISBN 978-958-761-204-2

Primera edición, 2004
Segunda edición, 2012

PREPARACIÓN EDITORIAL E IMPRESIÓN:
Editorial Universidad Nacional de Colombia
www.editorial.unal.edu.co
direditorial@unal.edu.co

Bogotá, Colombia

Prohibida la reproducción total o parcial por cualquier medio
sin la autorización escrita del titular de los derechos patrimoniales

Catalogación en la publicación Universidad Nacional de Colombia

Jiménez Moscoso, José Alfredo, 1973-
Algebra matricial con aplicaciones en estadística / José Alfredo Jiménez Moscoso.
-- Bogotá : Universidad Nacional de Colombia. Facultad de Ciencias, 2012
xiv, 478 p., il.

Incluye referencias bibliográficas

ISBN : 978-958-761-204-2

1. Matrices (Matemáticas) 2. Inversión de matrices 3. Estadística matemática
I. Tít.

CDD-21 512.9434 / 2012

A

Mi esposa
Mi hija
Mis padres

Contenido

Prólogo xiii

1 Preliminares **1**
1.1 Matrices 1
1.1.1 Conceptos básicos 2
1.1.2 Operaciones con matrices 3
1.1.3 Operaciones elementales sobre los renglones 7
1.1.4 Traza de una matriz 9
1.2 Inversa de una matriz 10
1.2.1 Método de Gauss-Jordan para calcular la inversa 10
1.3 Determinantes 12
1.3.1 Algunas fórmulas útiles para inversas 16
1.4 Tipos especiales de matrices cuadradas 16
1.5 Matrices particionadas 25
1.5.1 Definiciones y operaciones 25
1.5.2 Determinantes de matrices particionadas 34
1.5.3 Inversas de matrices particionadas 37
1.6 Espacio vectorial 43
1.6.1 Axiomas de un espacio vectorial 44
1.6.2 Bases 47
1.6.3 Espacios con producto interno 50
1.6.4 Complemento ortogonal 53
1.6.5 Subespacios asociados a una matriz 53
1.7 Sistemas de ecuaciones lineales 56
1.7.1 Método de eliminación de Gauss 57
1.8 Transformaciones lineales 59
1.8.1 Representación matricial de una transformación 59
1.9 Matrices con entradas complejas 61
1.9.1 Definición y propiedades básicas 61

1.9.2 Espacios vectoriales complejos 64
1.9.3 Solución de sistemas lineales con entradas complejas 66

2 Vectores característicos y valores característicos 69
2.1 Valores propios y vectores propios 70
2.1.1 Descomposición de Sylvester 80
2.2 Matrices semejantes y diagonalización 88
2.3 Valores propios complejos 98
2.4 Diagonalización de matrices simétricas 105
2.5 Vectores propios generalizados 115

3 Descomposición de matrices 127
3.1 Triangularización de una matriz 127
3.2 Factorización QR 143
3.3 Polinomio mínimo 148
3.4 Forma canónica de Jordan 153
3.5 Raíces cuadradas 162
3.5.1 Raíces cuadradas de matrices simétricas 180
3.5.2 Descomposición de Cholesky 183
3.6 Descomposición en valores singulares 185
3.6.1 Descomposición en valores singulares 188
3.6.2 Descomposición polar 192

4 Matrices complejas 197
4.1 Clases especiales de matrices complejas 197
4.1.1 Matrices hermitianas 197
4.1.2 Matrices antihermitianas 203
4.1.3 Matrices unitarias 205
4.1.4 Matrices normales 207
4.2 Factorizaciones 208
4.2.1 Forma canónica de Jordan 220
4.2.2 Descomposición en valores singulares 222
4.2.3 Descomposición polar 224

5 Formas bilineales 231
5.1 Formas bilineales 231
5.2 Formas cuadráticas 238
5.3 Diagonalización de una forma cuadrática 244
5.3.1 Diagonalización por completación de cuadrados 244
5.3.2 Diagonalización por transformación ortogonal 255
5.4 Ley de la inercia para formas cuadráticas 259

5.5 Clasificación de las formas cuadráticas 263
5.6 Aplicaciones a la geometría analítica 271
5.6.1 Rotación de ejes en $\mathbb{R}^2$ 277
5.6.1.1 Cambio de dirección de ejes en $\mathbb{R}^2$ conservando el mismo origen 278
5.6.2 Clasificación de las ecuaciones cuadráticas 283
5.6.3 Rotación de ejes en $\mathbb{R}^3$ 290
5.6.3.1 Cambio de dirección de ejes en $\mathbb{R}^3$ conservando el mismo origen 291
5.6.3.2 Fórmulas de Euler 291
5.6.4 Clasificación de las superficies cuádricas 296

6 Formas hermíticas 299
6.1 Forma hermítica 299
6.2 Forma cuadrática compleja 303
6.3 Diagonalización de una forma hermítica 305
6.4 Clasificación de formas cuadráticas complejas 310
6.5 Orden parcial entre matrices 311

7 Normas matriciales 315
7.1 Definición y resultados básicos 315
7.2 Tipos de normas matriciales 317
7.3 Condición de sistemas de ecuaciones lineales 324

8 Matrices idempotentes y productos especiales 333
8.1 Definición y propiedades 333
8.1.1 Factorización **QR** por reflexiones de Householder 338
8.2 Productos especiales 348

9 Inversa generalizada de matrices 363
9.1 Definición y propiedades básicas 363
9.2 Propiedades de las inversas generalizadas 367
9.3 Métodos para calcular inversas generalizadas 369
9.4 Vectores y valores propios 396
9.5 Solución de sistemas de ecuaciones lineales 398

10 Aplicaciones 409
10.1 Matrices estocásticas 409
10.2 Modelos genéticos 417
10.2.1 Herencia autosómica 418
10.2.2 Los cuadros de Punnett 421

10.3 Modelo de regresión lineal 425
10.3.1 Métodos de estimación de los parámetros del modelo 428
10.3.1.1 Método de mínimos cuadrados ordinarios 429
10.3.1.2 Forma operativa 431
10.3.1.3 Propiedades de los elementos de la matriz **H** 432
10.4 Multicolinealidad 433
10.4.1 Soluciones al problema de la multicolinealidad 434
10.4.1.1 Regresión por componentes principales 434
10.4.1.2 Propiedades de los componentes 436
10.5 Selección de carteras 441
10.5.1 Formulación matemática 441
10.5.2 Cartera con rentabilidad preestablecida 443
10.5.3 Cartera mínima con rentabilidad preestablecida 446

A Métodos iterativos para estimar valores propios y vectores propios 451
A.1 Valor propio y vector propio dominante 451
A.1.1 Método de la potencia 452

B Números complejos 459
B.1 Álgebra de los números complejos 459
B.1.1 Operaciones fundamentales 460
B.1.2 Representación polar 464

Bibliografía 469

Índice alfabético 473

Prólogo

El álgebra de matrices es en la actualidad un elemento esencial de los conocimientos matemáticos necesarios para ingenieros y científicos. Además, la comprensión de los métodos fundamentales del álgebra matricial es apropiada para sociólogos, economistas, estudiantes de pedagogía y de comercio.

A pesar de las diversas aplicaciones del álgebra matricial, la mayoría de textos de álgebra lineal no introducen estos temas, por eso en muchos casos no se encuentra un libro que se ajuste a los requerimientos y necesidades de ciertas materias. Estas notas de clase están basadas en el curso de álgebra matricial de la carrera de Estadística, las cuales han sido redactadas usando diferentes textos, resaltando principalmente los libros referenciados en la bibliografía como Apostol (1985), Asmar (1995), Barbolla & Sanz (1998), Bru et al. (2001), Graybill (1983), Schott (1997) y Searle (1982).

Este texto sirve de ayuda para aquellos estudiantes que toman diversas asignaturas en las cuales deben tener o les serían útiles los conocimientos del álgebra de matrices. Aunque en estas circunstancias siempre es inadecuado comenzar un curso de teoría de matrices, estas notas le permitirán al lector adquirir la práctica necesaria en el manejo de matrices.

El objetivo principal de estas notas consiste en capacitar al lector para que adquiera la habilidad de usar el álgebra de matrices en diferentes ámbitos, proporcionando conceptos como la diagonalización y factorización matricial, formas cuadráticas e inversas generalizadas de una manera sencilla; durante su desarrollo, se plantean ejemplos y ejercicios relacionados con la teoría.

Este material está escrito en forma secuencial, pues los contenidos previos son importantes para tener una mejor comprensión del desarrollo de

cada sección posterior, lo cual ayudará al lector a alcanzar su principal objetivo. Asimismo, proporciona un medio individual para estudiar el tema expuesto y es muy práctico como texto autodidáctico. Además, permitirá que el lector avance a su propio ritmo. De esta manera, este material puede ser usado por estudiantes con diferentes aptitudes, conocimientos y velocidades de lectura.

Espero que este material carezca de errores, sin embargo, "no importa el cuidado que se ponga, siempre se comete algún error" (Ley de Murphy). Por lo tanto, los errores que posea deseo conocerlos para poderlos corregir, en una próxima edición. Esta es tal vez la única forma de avanzar en un ambiente académico.

Agradezco la colaboración del Departamento de Matemáticas, que a través de su oficina de publicaciones me permitió la divulgación de este material. También quiero dar las gracias tanto a los colegas que evaluaron este manuscrito, en especial al profesor Leonardo Solanilla Chavarro, como a mis estudiantes del curso de Álgebra Matricial de la carrera de Estadística, por sus sugerencias y comentarios, los cuales fueron muy útiles en la redacción de este material. Adicionalmente, quiero agradecer al equipo editorial de la Universidad Nacional de Colombia, sede Bogotá, por la corrección de estilo.

José Alfredo Jiménez M.

Capítulo 1

Preliminares

Este capítulo es una recopilación de conceptos, procedimientos y resultados básicos que, por lo general, forman parte del primer curso de álgebra lineal. Por consiguiente, una gran parte de estos resultados aparecen sin prueba; además, en algunos casos se consideran temas que el lector debe manejar y que por su importancia son retomados posteriormente.

El propósito fundamental de este material es servir como prerrequisito para los siguientes capítulos y, como ya se mencionó, no se profundizará en los temas considerados en este capítulo. Si el lector tiene amplios conocimientos del contenido de este apartado, puede pasar de inmediato al siguiente capítulo, aunque es recomendable que desarrolle las Secciones 1.5 y 1.9.

1.1 Matrices

En esta sección se introducen los conceptos y las reglas básicas del álgebra de matrices. Entre los diferentes elementos estudiados por el álgebra lineal, uno de los más utilizados es el de *matriz*. Esto se debe a que la teoría de matrices ofrece, entre otras, la posibilidad de trabajar cómodamente con modelos de gran dimensión, tanto en número de variables, como de ecuaciones o datos, ya que brinda una notación simple y compacta para designar amplios conjuntos de información.

1.1.1 Conceptos básicos

En este apartado se presenta la definición formal del término *matriz*. Las matrices se denotan con letras mayúsculas y con minúsculas, los elementos que las constituyen.

Definición 1.1 *Una matriz A de tamaño $m \times n$ es un arreglo rectangular de $m \cdot n$ números reales (o complejos[1]) dispuestos en m filas y n columnas, escritos entre corchetes (o paréntesis), como sigue:*

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix},$$

donde los subíndices indican la "fila" y la "columna" de localización en la matriz de cada número real (o complejo). A los números $a_{11}, a_{12}, \dots, a_{mn}$ se les llama elementos o entradas de la matriz.

Observación

Cuando $m = n$, la matriz recibe el nombre de *cuadrada*; si es $m \neq n$, se denomina *rectangular*. Al conjunto de todas las matrices de tamaño $m \times n$ se le notará por $\mathcal{M}_{mn}$.

Definición 1.2 Matrices iguales

Sean las matrices reales $A = [a_{ij}]$ y $B = [b_{ij}]$, se dice que son iguales cuando teniendo el mismo tamaño, se verifica que

$$a_{ij} = b_{ij} \qquad \begin{array}{l} \forall\, i = 1, 2, \dots, m; \\ \forall\, j = 1, 2, \dots, n. \end{array}$$

[1] Si el lector no está familiarizado con estos números, puede consultar el Apéndice B.

1.1.2 Operaciones con matrices

En esta sección se consideran las operaciones con matrices; además, se recuerda que solo se pueden sumar matrices que tienen el mismo tamaño, y para que el producto sea posible, es preciso que el número de columnas de la primera matriz coincida con el número de filas de la segunda.

Definición 1.3 *Dadas $A = [a_{ij}]$ y $B = [b_{ij}]$, matrices de tamaño $m \times n$, la suma $A + B$ es una matriz $C = [c_{ij}]$ de tamaño $m \times n$, donde*

$$c_{ij} = a_{ij} + b_{ij} \qquad \forall\, i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n.$$

Teorema 1.1 Propiedades básicas de la suma

Para todas A, B y C matrices de tamaño $m \times n$, se verifica que

1 *Conmutativa: $A + B = B + A$.*

2 *Asociativa: $\big(A + B\big) + C = A + \big(B + C\big)$.*

3 *Existencia de elemento neutro o matriz nula: Existe una matriz* O *de tamaño $m \times n$ en donde todos sus elementos son iguales a cero, tal que $\forall A$ de tamaño $m \times n$, se verifica que*

$$A + \mathrm{O} = \mathrm{O} + A = A.$$

4 *Elemento opuesto: Para toda matriz A de tamaño $m \times n$, existe una matriz que llamaremos matriz opuesta de A y denotaremos por $-A$, que verifica*

$$A + \big(-A\big) = \mathrm{O}\,.$$

La última propiedad permite, dadas dos matrices $A = [a_{ij}]$ y $B = [b_{ij}]$ del mismo tamaño $m \times n$, introducir el concepto de matriz diferencia $A - B$, la cual puede ser definida como sigue:

$$A - B = A + \big(-B\big).$$

Demostración.

Sean $A = [a_{ij}]$, $B = [b_{ij}]$ y $C = [c_{ij}]$.

1. $a_{ij} + b_{ij} = b_{ij} + a_{ij}$.
2. $\big(a_{ij} + b_{ij}\big) + c_{ij} = a_{ij} + \big(b_{ij} + c_{ij}\big)$.
3. Es evidente que, para toda A de tamaño $m \times n$, se verifica que
$$a_{ij} + 0 = 0 + a_{ij} = a_{ij}.$$
4. Al tomar $-A = [-a_{ij}]$, se verifica que
$$a_{ij} + \big(-a_{ij}\big) = \big(-a_{ij}\big) + a_{ij} = 0. \quad \blacksquare$$

Definición 1.4 *El producto de una matriz $A = [a_{ij}]$ de tamaño $m \times n$ por un escalar $\alpha \in \mathbb{R}$ es una matriz $C = [c_{ij}]$ del mismo tamaño que A, de elementos*

$$c_{ij} = \alpha a_{ij} \qquad \forall\, i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n,$$

esto es, los elementos de C se obtienen multiplicando los elementos correspondientes de A por α.

El resultado de efectuar el producto de una matriz A por un escalar α se simboliza por αA y se lee *multiplicación de A por α*.

Teorema 1.2 Propiedades de la multiplicación por un escalar

Para todas A y B de tamaño $m \times n$ y $\alpha, \beta \in \mathbb{R}$, se satisface que

a) $\alpha\big(A + B\big) = \alpha A + \alpha B$, *b)* $\big(\alpha + \beta\big)A = \alpha A + \beta A$,

c) $\alpha\big(\beta A\big) = \big(\alpha\beta\big)A$, *d)* $1A = A$.

Demostración.
Sean $A = [a_{ij}]$ y $B = [b_{ij}]$, entonces

a) $\alpha\big(A + B\big) = \alpha\,[a_{ij} + b_{ij}] = \alpha\,[a_{ij}] + \alpha\,[b_{ij}]\,.$

b) $\big(\alpha + \beta\big)A = \big(\alpha + \beta\big)\,[a_{ij}] = \alpha\,[a_{ij}] + \beta\,[a_{ij}]\,.$

c) $\alpha\big(\beta A\big) = \alpha\big(\beta\,[a_{ij}]\big) = \big(\alpha\beta\big)\,[a_{ij}]\,.$

d) $1A = 1\,[a_{ij}] = [a_{ij}]\,.$ ■

Definición 1.5 Matriz identidad

Una matriz $A = [a_{ij}]$ de tamaño $n \times n$ cuyos elementos son

$$a_{ij} = \begin{cases} 1 & si\ i = j \\ 0 & si\ i \neq j \end{cases}$$

se llama matriz identidad y se denota por I_n.

Definición 1.6 *Sean $A = [a_{ij}]$ y $B = [b_{jk}]$ matrices de tamaño $m \times n$ y $n \times p$, respectivamente. Entonces el producto de las matrices A y B, operación que se denotará por $A.B$, es una matriz C de tamaño $m \times p$, cuyo elemento genérico c_{ik} $(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, p)$ es*

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}.$$

Teorema 1.3 Propiedades del producto de matrices

Sean A, B, C y D matrices reales tales que $A \in \mathcal{M}_{mn}$, $B, C \in \mathcal{M}_{np}$ y $D \in \mathcal{M}_{pq}$. Entonces se satisface que

1. *Asociativa: $A.\big(B.D\big) = \big(A.B\big).D$.*

2. *Distributiva:*

$$a)\ A.\big(B + C\big) = A.B + A.C \qquad b)\ \big(B + C\big).D = B.D + C.D.$$

3. *El producto por una matriz nula* O *del tamaño adecuado es una matriz nula.*

4. *En general, esta operación matricial no es conmutativa:*

$$A.B \neq B.A.$$

5. *Existen matrices* I_m *e* I_n *tales que*

$$I_m.A = A \qquad y \qquad A.I_n = A.$$

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.7 Transpuesta

Si $A = [a_{ij}]$ *es una matriz real de tamaño* $m \times n$*, se llama transpuesta de* A *a la matriz* $B = [b_{ij}]$ *de tamaño* $n \times m$ *cuyo elemento* $b_{ij} = a_{ji}$*. Se denota por* A^t*.*

Teorema 1.4 Propiedades de la transpuesta

Sean A *y* B *matrices de tamaño* $m \times n$*,* C *una matriz de tamaño* $n \times m$ *y sea* $\alpha \in \mathbb{R}$*. Entonces*

1. $\left(A^t\right)^t = A.$
2. $\left(A \pm B\right)^t = A^t \pm B^t.$
3. $\left(\alpha A\right)^t = \alpha A^t.$
4. $\left(AC\right)^t = C^t A^t.$

Demostración.
Queda como ejercicio para el lector. ■

1.1.3 Operaciones elementales sobre los renglones

Sea A una matriz real de tamaño $m \times n$, entonces las operaciones elementales en las filas de la matriz son:

R_1 Multiplicar cada elemento de la i-ésima fila por un escalar $\alpha \neq 0$.

R_2 Sumar a la i-ésima fila un múltiplo de la k-ésima fila.

R_3 Intercambiar (permutar) dos filas.

Y las operaciones elementales en las columnas de la matriz son:

C_1 Multiplicar cada elemento de la j-ésima columna por un escalar $\alpha \neq 0$.

C_2 Sumar a la j-ésima columna un múltiplo de la l-ésima columna.

C_3 Intercambiar (permutar) dos columnas.

Definición 1.8 Matrices elementales

Una matriz $E_{kl}(\alpha)$ de tamaño $m \times m$ se llama matriz elemental si es el resultado de aplicar una operación elemental a la matriz identidad I_m.

Realizar una operación elemental en una fila (o columna) de una matriz A es equivalente a premultiplicar (o multiplicar) a A, respectivamente, por la matriz elemental adecuada. Esto se tiene de la definición de multiplicación de matrices, la cual nos aclaró el hecho de que premultiplicar (o multiplicar) una matriz A por una matriz elemental daba el mismo resultado que aplicar la operación elemental a la fila correspondiente de la matriz A.

Notación

La notación que se usará para los tres tipos de operaciones R_1, R_2 y R_3 con matrices elementales es la siguiente:

- La matriz elemental tipo R_1 es una matriz $E_{kl}(\alpha) = [\nu_{ij}]$, cuyos elementos son

$$\nu_{ij} \underset{k=l}{=} \begin{cases} 1 & \text{si } i = j \neq k, \\ \alpha & \text{si } i = j = k, \\ 0 & \text{si } i \neq j. \end{cases}$$

  Nótese que es una matriz diagonal.

- La matriz elemental tipo R_2 es una matriz $E_{kl}(\alpha) = [\nu_{ij}]$:

$$\nu_{ij} \underset{k \neq l}{=} \begin{cases} 1 & \text{si } i = j, \\ \alpha & \text{si } i = k, j = l, \\ 0 & \text{en otro caso.} \end{cases}$$

  Esta matriz es triangular superior (inferior) dependiendo de la relación de orden que exista entre r y s. Además, si $k = l$ coincide con la matriz elemental tipo R_1.

- Matriz elemental tipo R_3 es una matriz $E_{kl}(1) = [\nu_{ij}]$:

$$\nu_{ij} \underset{\alpha=1}{=} \begin{cases} 1 & \text{si } i = j, i \neq k, i \neq l, \\ 1 & \text{si } i = k, j = l, \\ 1 & \text{si } i = l, j = k, \\ 0 & \text{en otro caso.} \end{cases}$$

Definición 1.9 Matriz escalonada

Se dice que una matriz es escalonada si el número de ceros que precede al primer elemento diferente de cero de una fila aumenta fila por fila hasta tener posiblemente filas de solo ceros.

Definición 1.10 Forma escalonada reducida

Una matriz se dice que es escalonada reducida si verifica las siguientes condiciones:

i. Es una matriz escalonada.

ii. El primer elemento no nulo (por la izquierda) de cada fila no nula es un 1 y este es el único elemento diferente de cero que se encuentra en la respectiva columna.

iii. Las filas nulas, si existen, están en la parte inferior de la matriz.

1.1.4 Traza de una matriz

En esta sección se estudiará una característica de las matrices cuadradas, la cual se expresa a través de un número llamado *traza*.

Definición 1.11 Traza de una matriz

Sea $A = [a_{ij}]$ una matriz real de tamaño $n \times n$, la suma de los elementos de la diagonal principal se llama traza de A y se denota como $\operatorname{tr}(A)$*, o sea*

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}. \tag{1.1}$$

Teorema 1.5 Propiedades

1. $\operatorname{tr}(I_n) = n$, *siendo I_n la matriz identidad de tamaño $n \times n$.*

2. $\operatorname{tr}(\mathrm{O}) = 0$ *siendo* O *la matriz nula de tamaño $n \times n$.*

3. $\operatorname{tr}(A) = \operatorname{tr}(A^t)$.

4. $\operatorname{tr}(A.A^t) = \operatorname{tr}(A^t.A) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2$.

5. $\operatorname{tr}(\alpha A) = \alpha \operatorname{tr}(A)$, *con* $\alpha \in \mathbb{R}$.

6. *Si A y B son del mismo tamaño,* $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.

7. *Si son posibles los productos $A.B$ y $B.A$, entonces se verifica*

$$\operatorname{tr}(A.B) = \operatorname{tr}(B.A).$$

8. $\operatorname{tr}(A.X) = 0$, *para toda $X \in \mathcal{M}_{nn}$, implica que* $A = \mathrm{O}$.

Demostración.

Queda como ejercicio para el lector. ■

1.2 Inversa de una matriz

Es sabido que todo número $\alpha \neq 0$ tiene un inverso α^{-1} tal que

$$\alpha\alpha^{-1} = \alpha^{-1}\alpha = 1. \tag{1.2}$$

Este hecho permite resolver las ecuaciones del tipo $\alpha x = \beta$, ya que multiplicando por α^{-1} se obtiene $x = \alpha^{-1}\beta$.

En este apartado se define un tipo de matriz que tiene una propiedad análoga en la teoría de matrices, la *matriz inversa*.

Definición 1.12 Inversa de una matriz

Sea A una matriz real de tamaño $n \times n$, si existe una matriz real B de tamaño $n \times n$ tal que

$$A.B = B.A = I_n. \tag{1.3}$$

Entonces B se denota por A^{-1} y recibe el nombre de matriz inversa.

Notación

Si A tiene inversa, entonces A se llama *matriz no singular o invertible*; en cambio, si A no tiene inversa, entonces A se llama *matriz singular o no invertible*.

1.2.1 Método de Gauss-Jordan para calcular la inversa

Para encontrar la inversa de una matriz cuadrada A de tamaño $n \times n$, se procede de la siguiente manera:

1. Se forma la matriz aumentada $B = \big(A \mid I_n\big)$ de tamaño $n \times 2n$.
2. Se aplican operaciones elementales entre filas hasta llevar a B a una matriz escalonada reducida $C = \big(\tilde{A}_1 \mid \tilde{A}_2\big)$.
3. Se decide si A es no singular.
   a. Si $\tilde{A}_1 = I_n$, entonces $\tilde{A}_2 = A^{-1}$.
   b. Si $\tilde{A}_1 \neq I_n$, entonces $\tilde{A}_1$ tiene una fila de ceros. En este caso, A es singular, es decir A^{-1} no existe.

Teorema 1.6 Propiedades de la inversa de una matriz

1. *Si una matriz A tiene inversa, esta es única.*

2. *La inversa del producto de un escalar no nulo por una matriz es el producto del inverso multiplicativo del escalar por la inversa de la matriz. En símbolos,*

$$\big(\alpha A\big)^{-1} = \frac{1}{\alpha} A^{-1}, \qquad con \qquad \alpha \in \mathbb{R},\, \alpha \neq 0.$$

3. *La inversa de la inversa es la matriz original. En símbolos,*

$$\big(A^{-1}\big)^{-1} = A.$$

4. *La inversa de una matriz transpuesta es la transpuesta de la inversa. En símbolos,*

$$\big(A^t\big)^{-1} = \big(A^{-1}\big)^t.$$

5. *Si A y B son dos matrices invertibles y del mismo tamaño, el producto $A.B$ es invertible y, además,*

$$a)\ \big(A.B\big)^{-1} = B^{-1}.A^{-1},$$
$$b)\ \Big[\big(A.B\big)^{-1}\Big]^t = \big(A^t\big)^{-1}\big(B^t\big)^{-1}.$$

6. *Si A es una matriz invertible,*

$$a)\ B.A = \mathrm{O} \qquad \Rightarrow \qquad B = \mathrm{O}\,.$$
$$b)\ B.A = C.A \qquad \Rightarrow \qquad B = C.$$

Demostración.
Queda como ejercicio para el lector. ■

1.3 Determinantes

Los *determinantes* permiten determinar cuando una matriz cuadrada es invertible. Un determinante de n-ésimo orden es una expresión asociada con una matriz $A = [a_{ij}]$ de tamaño $n \times n$, como se explica a continuación empezando con $n = 2$.

Definición 1.13 *Sea $A = [a_{ij}]$ una matriz de tamaño 2×2. Entonces, el determinante de A se define por*

$$\det A = a_{11}.a_{22} - a_{12}.a_{21}. \tag{1.4}$$

Con frecuencia, se denotará el $\det A$ por

$$|A| \qquad \text{o} \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix};$$

aquí se usan barras (mientras que una matriz tiene corchetes).

Definición 1.14 *Sea $A = [a_{ij}]$ una matriz de tamaño 3×3. Entonces, el determinante de A se puede escribir en términos de los determinantes de matrices 2×2, como sigue:*

$$\det A = a_{11}. \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}. \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}. \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \tag{1.5}$$

o en la forma explícita siguiente:

$$|A| = a_{11} \left[a_{22}a_{33} - a_{23}a_{32}\right] - a_{12} \left[a_{21}a_{33} - a_{23}a_{31}\right] + a_{13} \left[a_{21}a_{32} - a_{22}a_{31}\right].$$

También hay un método para memorizar esta fórmula, llamado *Regla de Sarrus*, que consiste en agregar las dos primeras columnas a la derecha de A, y se suman todos los productos de los elementos que van de la izquierda superior a la derecha inferior y se restan todos los productos de los elementos que van de la izquierda inferior a la derecha superior. Así:

$$|A| = \begin{vmatrix} a_{11} & \searrow & a_{12} & \swarrow & a_{13} \\ a_{21} & & a_{22} & & a_{23} \\ a_{31} & \nearrow & a_{32} & \swarrow & a_{33} \end{vmatrix} \begin{matrix} \swarrow & a_{11} & \nearrow & a_{12} \\ & a_{21} & & a_{22} \\ \swarrow & a_{31} & \searrow & a_{32} \end{matrix}$$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}.$$

Nota 1.1 La Regla de Sarrus no se puede aplicar para calcular determinantes de matrices de tamaño 4×4, 5×5 o para matrices de más alto orden.

Hasta ahora, se ha evaluado los determinantes para matrices de tamaño 2×2 y 3×3. El determinante de una matriz de tamaño $n \times n$ se obtiene usando un método en el cual se reduce el problema a la evaluación de determinantes de matrices de orden $n - 1$; el proceso se repite sucesivamente hasta llegar a las matrices de tamaño 2×2. Nótese que este procedimiento fue empleado para calcular el determinante en la expresión (1.5), eliminando de A la fila y la columna, que indican el primer y segundo subíndice del elemento a_{ij} por el que van multiplicados, los determinantes de las submatrices de tamaño 2×2, los cuales reciben el nombre de menores, y cuando se les asocia los signos $+, -, +$ se denominan *cofactores* o *adjuntos*. Las definiciones de estos conceptos son:

Definición 1.15 Menor y cofactor

Sea $A = [a_{ij}]$ una matriz real de tamaño $n \times n$.

1. *Se le llama menor complementario (i, j), al determinante de la submatriz de tamaño $\big(n-1\big)\times\big(n-1\big)$, que resulta de suprimir la i-ésima fila y la j-ésima columna de A y se denota por $M_{ij}(A)$.*

2. *El adjunto o cofactor (i, j) de A viene dado por*

$$C_{ij}(A) = \big(\text{-}1\big)^{i+j} M_{ij}(A) \quad \textit{donde} \quad \big(\text{-}1\big)^{i+j} = \begin{cases} 1 & \textit{si } i+j \textit{ es par;} \\ \text{-}1 & \textit{si } i+j \textit{ es impar.} \end{cases}$$

Definición 1.16 Matriz de cofactores

La matriz $C = [C_{ij}(A)]$, donde el elemento $C_{ij}(A)$ es el cofactor (i, j) de A, se denomina matriz de cofactores.

Teorema 1.7 Fórmula o expansión de Laplace

Sea $A = [a_{ij}]$ una matriz de tamaño $n \times n$. Entonces, el determinante de A se puede desarrollar usando:

i) La expansión de Laplace por la i-ésima fila como

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij}(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}(A).$$

ii) La expansión de Laplace por la j-ésima columna como

$$\det A = \sum_{i=1}^{n} a_{ij} C_{ij}(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}(A).$$

Demostración.
Queda como ejercicio para el lector. ■

Teorema 1.8 Propiedades de los determinantes

Dadas A, B y C matrices de tamaño $n \times n$ y $\alpha \in \mathbb{R}$, se verifica que

i) $\det A^t = \det A$.

ii) Si se multiplica solo una fila (o columna) de la matriz A por un escalar α, entonces el determinante queda multiplicado por α.

iii) El determinante de la matriz αA es

$$\det(\alpha A) = \alpha^n \det A.$$

iv) Si todos los elementos de una fila (o columna) de A son cero, el valor del determinante es cero.

v) Si las matrices A, B y C difieren exclusivamente en los elementos de la j-ésima columna, siendo los elementos de esta columna para la matriz C la suma de los respectivos elementos de la j-ésima columna de las matrices A y B, entonces

$$\det C = \det A + \det B.$$

El mismo resultado se cumple cuando las tres matrices difieren de manera análoga en una fila.

vi) Si dos filas (o columnas) cualesquiera de A se intercambian, el valor del determinante se multiplica por -1.

vii) Si dos filas (o columnas) de A son proporcionales o iguales, el valor del determinante es cero.

viii) Si se suma un múltiplo escalar de una fila (o columna) de A a otra fila (o columna) de A, entonces el determinante no cambia.

ix) a) $\det\big(A.B\big) = \det A.\det B.$

b) Para cualquier $k \in \mathbb{N}$, $k \neq 0$ $\det\big(A^k\big) = \big(\det A\big)^k$.

c) Si A es invertible, entonces $\det\big(A^{-1}\big) = \big(\det A\big)^{-1}$.

Demostración.
Queda como ejercicio para el lector. ■

Teorema 1.9 *Sea A una matriz de tamaño $n \times n$, A es invertible si y solo si* $\det A \neq 0$.

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.17 Matriz adjunta

Sea A una matriz de tamaño $n \times n$. La matriz transpuesta de la matriz de cofactores $C_{ij}(A)$ es la adjunta de A y se representa por $\mathrm{Adj}(A)$:

$$\mathrm{Adj}(A) = \begin{bmatrix} C_{11}(A) & C_{21}(A) & \dots & C_{n1}(A) \\ C_{12}(A) & C_{22}(A) & \dots & C_{n2}(A) \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n}(A) & C_{2n}(A) & \dots & C_{nn}(A) \end{bmatrix}. \tag{1.6}$$

Teorema 1.10 *Sea A una matriz de tamaño $n \times n$, si* $\det A \neq 0$*, entonces*

$$A^{-1} = \mathrm{Adj}(A)/\det A.$$

Demostración.
Queda como ejercicio para el lector. ■

1.3.1 Algunas fórmulas útiles para inversas

- Para una matriz invertible de tamaño 2×2, se obtiene

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \tag{1.7}$$

- Para una matriz invertible de tamaño 3×3, se obtiene

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} C_{11}(A) & C_{21}(A) & C_{31}(A) \\ C_{12}(A) & C_{22}(A) & C_{32}(A) \\ C_{13}(A) & C_{23}(A) & C_{33}(A) \end{bmatrix} \tag{1.8}$$

1.4 Tipos especiales de matrices cuadradas

Los tipos de matrices que se analizan a continuación tienen características particulares y, como se presentan frecuentemente en el desarrollo de la teoría y en las aplicaciones, han recibido denominaciones especiales.

Definición 1.18 Matrices triangulares

Una matriz cuadrada real $A = [a_{ij}]$ cuyos elementos abajo de la diagonal principal son todos cero, es decir $a_{ij} = 0$ para $i > j$, se llama matriz triangular superior. De manera análoga, una matriz triangular inferior es una matriz cuadrada real A cuyos elementos arriba de la diagonal principal son cero, es decir $a_{ij} = 0$ para $i < j$.

Teorema 1.11 Propiedades de las matrices triangulares

Sean A, $B \in \mathcal{M}_{nn}$ matrices triangulares superiores (inferiores) y $\alpha \in \mathbb{R}$, entonces:

i) Las matrices $A + B$ y αA son triangulares superiores (inferiores).

ii) La matriz $A.B$ es también triangular superior (inferior).

iii) El $\det(A)$ es igual al producto de los elementos de la diagonal principal.

iv) La transpuesta de A es triangular inferior (superior).

v) La matriz A es no singular si y solo si cada uno de los elementos de la diagonal es distinto de cero.

vi) Si A es invertible, entonces A^{-1} es triangular superior (inferior).

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.19 Matrices simétricas

Una matriz cuadrada real $A = [a_{ij}]$ se llama simétrica si la transposición la mantiene invariable, es decir $[a_{ij}] = [a_{ji}]$.

Teorema 1.12 Propiedades de las matrices simétricas

Sean A y B matrices simétricas de tamaño $n \times n$ y $\alpha \in \mathbb{R}$, entonces:

i) $A + B$ *y* αA *son simétricas.*

ii) Cuando $A.B = B.A$*, entonces* $A.B$ *es simétrica. Sin embargo, esto no es cierto si* A *y* B *no conmutan en el producto.*

iii) Si A *es invertible entonces su inversa* A^{-1} *también es simétrica.*

iv) Dada una matriz cualquiera C *de tamaño* $m \times n$*,*

a) Si $m = n$*, la matriz* $\frac{1}{2}\left(C + C^t\right)$ *es simétrica.*

b) Si $m \neq n$ *o si* $m = n$*, las matrices* $\left(C.C^t\right)$ *y* $\left(C^t.C\right)$ *son simétricas.*

Demostración.
Queda como ejercicio para el lector. ■

Un tipo especial de matrices simétricas son las *matrices escalares.*

Definición 1.20 Matriz escalar

Una matriz real S *de tamaño* $n \times n$ *se llama matriz escalar si resulta de la multiplicación de* I_n *por un escalar* $c \in \mathbb{R}$*, es decir*

$$S = c.I_n.$$

Teorema 1.13 Propiedades de las matrices escalares

Si $A = aI_n$ *y* $B = bI_n$ *son matrices de tamaño* $n \times n$ *escalares,* $a, b \in \mathbb{R}$*, entonces:*

i) Las matrices $A + B$ *y* $A.B$ *son también matrices escalares.*

ii) *El* $\det(A) = a^n$.

iii) *La transpuesta de* A *es una matriz escalar.*

iv) *La matriz* A *es no singular si y solo si* $a \neq 0$.

v) *Si* A *es invertible, entonces* A^{-1} *es también una matriz escalar.*

vi) *Si* C *es una matriz de tamaño* $n \times n$*, entonces* C *conmuta con* A*, es decir,* $CA = AC$.

Demostración.
Queda como ejercicio para el lector. ■

Cuando todos los elementos s_{ii} de S no son iguales al escalar c, se tiene un nuevo tipo de matrices simétricas: las *matrices diagonales.*

Definición 1.21 Matrices diagonales

Una matriz cuadrada real $A = [a_{ij}]$ *cuyos elementos arriba y abajo de la diagonal principal son todos cero, es decir que* $a_{ij} = 0$ *para toda* $i \neq j$*, se llama matriz diagonal.*

Teorema 1.14 Propiedades de las matrices diagonales

Si $D = [d_{ii}]$ *es una matriz diagonal de tamaño* $n \times n$*, entonces:*

i) *Su producto por otra matriz diagonal también corresponde a una matriz diagonal.*

ii) *El* $\det D$ *es igual al producto de los elementos de la diagonal principal.*

iii) D *es una matriz no singular si y solo si todos los elementos de la diagonal son distintos de cero.*

iv) Si sus elementos de la diagonal principal $d_{11}, d_{22}, \ldots, d_{nn}$ *son todos distintos de cero,* D^{-1} *es también una matriz diagonal con elementos en la diagonal principal iguales a* $1/d_{11}, 1/d_{22}, \ldots, 1/d_{nn}$.

Demostración.

i) Sea $C = [c_{ij}]$. Entonces, el elemento ij de DC es

$$\sum_{k=1}^{n} d_{ik} c_{kj},$$

pero como D y C son matrices diagonales, entonces $d_{ik} = 0$ si $i \neq k$ y $c_{kj} = 0$ si $k \neq j$.

Luego, el término $d_{ik}c_{kj} = 0$ si $i \neq j$. Por lo tanto, el único término que posiblemente es distinto de cero en esta suma es cuando $i = j = k$, es decir el término $d_{jj}c_{jj}$ que corresponde al elemento jj de $D.C$.

ii) Se procede por inducción.

Si se desarrolla el $\det D = |D|$ por la primera columna, se obtiene $|D| = d_{11}|D'|$, donde D' es una submatriz real de tamaño $(n-1) \times (n-1)$ obtenida al borrar la primera fila y la primera columna de D. Ahora, obsérvese que D' también es diagonal. Se deja al lector completar los detalles de la prueba.

iii) Dado que una matriz real de tamaño $n \times n$ es no singular si y solo si su determinante es diferente de cero, si D es diagonal, de la parte *ii)* se tiene que $|D| = d_{11} \cdot d_{22} \cdot \ldots \cdot d_{nn}$, y este es distinto de cero si y solo si cada $d_{ii} \neq 0$.

iv) Esta se sigue inmediatamente a partir de *i)* y de *iii)*. Si cada $d_{ii} \neq 0$,

$$\begin{bmatrix} d_{11} & 0 & \ldots & 0 \\ 0 & d_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{nn} \end{bmatrix} \begin{bmatrix} 1/d_{11} & 0 & \ldots & 0 \\ 0 & 1/d_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1/d_{nn} \end{bmatrix} = I_n. \quad \blacksquare$$

Definición 1.22 Matrices antisimétricas

Una matriz cuadrada real $A = [a_{ij}]$ se llama antisimétrica si la transposición da como resultado la negativa de A, es decir $A^t = -A$.

Teorema 1.15 Propiedades de las matrices antisimétricas

Sean A y B matrices cuadradas antisimétricas y $\alpha \in \mathbb{R}$, entonces:

i) $A + B$ y αA son antisimétricas.

ii) Cuando $AB = BA$, entonces AB es antisimétrica. Sin embargo, esto no es cierto si A y B no conmutan en el producto.

iii) Si A es invertible, entonces su inversa A^{-1} también es antisimétrica.

iv) Dada una matriz cualquiera C de tamaño $n \times n$, entonces la matriz $\frac{1}{2}\left(C - C^t\right)$ es antisimétrica.

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.23 Matriz ortogonal

Sea $A = [\vec{a}_1\ \vec{a}_2\ \ldots\ \vec{a}_n]$ una matriz real de tamaño $n \times n$, donde $\vec{a}_i$ es un vector $n \times 1$ formado con los elementos de la i-ésima columna de A. Entonces, A es ortogonal si y solo si

$$\vec{a}_i^t \vec{a}_j = \vec{a}_i \cdot \vec{a}_j = \begin{cases} 1 & si \quad i = j, \\ 0 & si \quad i \neq j. \end{cases}$$

Este producto en algunos casos se denota con el símbolo δ_{ij}, el cual se lee delta $i\,j$ y se denomina Delta de Kronecker.

Teorema 1.16 Propiedades de las matrices ortogonales

Sean A y B matrices ortogonales de tamaño $n \times n$ y $\alpha \in \mathbb{R}$, entonces:

i) *A es ortogonal si y solo si $A^t = A^{-1}$ o idénticamente si y solo si*

$$A^t A = I_n. \tag{1.9}$$

ii) *AB y BA son ortogonales pero, en general, $A + B$ y αA no lo son.*

iii) *El valor absoluto del* $\det A$ *es 1.*

iv) *La transpuesta de una matriz ortogonal es ortogonal.*

v) *Dada una matriz antisimétrica C de tamaño $n \times n$, entonces la matriz $A = \big(I_n - C\big)\big(I_n + C\big)^{-1}$ es ortogonal.*

Demostración.

i) Por la unicidad de la inversa de una matriz y sus propiedades, se tiene que

$$A.A^t = I_n \Rightarrow A^{-1}.A.A^t = A^{-1} \Rightarrow A^t = A^{-1}$$
$$A^t = A^{-1} \Rightarrow A.A^t = A.A^{-1} \Rightarrow A.A^t = I_n.$$

ii) Si A y B son ortogonales, entonces $A.B$ también lo es, ya que

$$\big(A.B\big)\big(A.B\big)^t = A.B.B^t.A^t = A.I_n.A^t = A.A^t = I_n.$$

Análogamente, se prueba para $B.A$.

iii) Si A es ortogonal, como $A^t.A = I_n$, se tiene que

$$\det\big(I_n\big) = \det\big(A^t.A\big) = \det\big(A^t\big)\det A = 1$$

y como $\det\big(A^t\big) = \det A$, se ve que $\big(\det A\big)^2 = 1$, y por tanto,

$$\det A = \pm 1.$$

iv) Obsérvese que $A^t.A = I_n$ se puede escribir como $A^t\big(A^t\big)^t = I_n$.

v) Si $C \in \mathcal{M}_{nn}$ es antisimétrica, se tiene que $C^t = -C$, y por las propiedades de la matriz transpuesta, resulta que

$$\begin{aligned} A^tA &= \Big[\big(I_n - C\big)\big(I_n + C\big)^{-1}\Big]^t \big(I_n - C\big)\big(I_n + C\big)^{-1} \\ &= \Big[\big(I_n + C\big)^{-1}\Big]^t \big(I_n - C\big)^t\big(I_n - C\big)\big(I_n + C\big)^{-1} \\ &= \big(I_n - C\big)^{-1}\big(I_n + C\big)\big(I_n - C\big)\big(I_n + C\big)^{-1} = I_n, \end{aligned}$$

porque $\big(I_n + C\big)\big(I_n - C\big) = \big(I_n - C\big)\big(I_n + C\big)$.

Así pues, A es ortogonal en virtud de la ecuación (1.9). ■

Definición 1.24 *Una matriz ortogonal A tal que* $\det A = 1$ *se llama matriz ortogonal propia y si el* $\det A = -1$*, se denomina matriz ortogonal impropia.*

Ejemplo 1.1 *¿Es ortogonal la matriz* $A = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}$ *?*

Solución.

Al multiplicar a A por la derecha por A^t, se obtiene

$$AA^t = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Esto muestra, por la ecuación (1.9), que A es ortogonal.

Definición 1.25 Matriz de reflexión

Una matriz ortogonal A tal que $A^2 = I_n$ se llama matriz de reflexión.

Definición 1.26 Matriz de permutación

La matriz elemental tipo R_3 de tamaño $n \times n$ se denomina matriz de permutación, ya que resulta de intercambiar (permutar) el orden de las filas de la matriz I_n.

Teorema 1.17 *Sea P una matriz de permutación, entonces:*

a) Para cualquier matriz A, se puede obtener PA a partir de A permutando las filas de A exactamente como se permutaron las filas de I_n para obtener P.

b) P es no singular y ortogonal.

Demostración.

a) Esto se sigue fácilmente de las definiciones de la multiplicación de matrices y de matrices de permutación.

b) Separe P en sus respectivas filas $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_n$, que son tan solo las filas ${\vec{e}_i}^t$ de I_n en cierto orden. Entonces P^t tiene como columnas a ${\vec{r}_i}^t$. La definición de la multiplicación de matrices implica que el elemento (i, j) de PP^t es simplemente $\vec{r}_i {\vec{r}_j}^t$, y esto es

$$\vec{r}_i {\vec{r}_j}^t = \delta_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Es decir, $PP^t = I_n$. De manera análoga, en términos de las columnas de P, se demuestra que $P^t P = I_n$. ■

Definición 1.27 *Una matriz N de tamaño $n \times n$ se dice que es nilpotente de índice k, si $N^k = \mathrm{O}$ para algún $k \in \mathbb{N}$, pero $N^{k-1} \neq \mathrm{O}$.*

Teorema 1.18 *Si N es una matriz cuadrada nilpotente, entonces:*

$$\det N = 0 \qquad y \qquad \operatorname{tr}(N) = 0.$$

Demostración.

Queda como ejercicio para el lector. ■

1.5 Matrices particionadas

Una característica importante en el trabajo que se hace con matrices reales es el de considerar la matriz A como una lista de vectores columna en lugar de simplemente una serie rectangular de números. Este método ha sido tan útil que ahora se desea considerar otras particiones de la matriz A, tomando por regla la división de A tanto horizontal como verticalmente. En esta sección se estudia la forma de particionar una matriz en submatrices que nos permitan desarrollar de manera más sencilla las mismas operaciones que definimos anteriormente para las matrices.

1.5.1 Definiciones y operaciones

Si $A = [a_{ij}]$ es la matriz que se obtiene después de que algunas filas y/o columnas de A se han eliminado es llamada una *submatriz* de A. Con frecuencia es conveniente particionar una matriz en submatrices y considerarla como una matriz cuyos elementos son estas submatrices.

Definición 1.28 *Una matriz A de tamaño $m \times n$ puede particionarse de la siguiente manera*

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{bmatrix}, \tag{1.10}$$

donde A_{11} es la submatriz real de tamaño $m_1 \times n_1$ formada por los elementos de A que ocupan las m_1 primeras filas y las n_1 primeras columnas; A_{12} es la submatriz real de tamaño $m_1 \times n_2$ formada por los elementos de A que ocupan las m_1 primeras filas y las columnas $n_1 + 1, \dots, n_1 + n_2$, y así sucesivamente. En general, A_{ij} es la submatriz real de tamaño $m_i \times n_j$ formada por los elementos de A que ocupan las filas

$$m_1 + \ldots + m_{i-1} + 1, m_1 + \ldots + m_{i-1} + 2, \ldots, m_1 + \ldots + m_{i-1} + m_i$$

y las columnas

$$n_1 + \ldots + n_{j-1} + 1, n_1 + \ldots + n_{j-1} + 2, \ldots, n_1 + \ldots + n_{j-1} + n_j,$$

siendo m_i y n_j números naturales tales que

$$m = \sum_{i=1}^{r} m_i \qquad y \qquad n = \sum_{j=1}^{s} n_j.$$

Se denotará el tamaño de la partición de A por bloques de la siguiente manera:

$$(m_1 + m_2 + \ldots + m_r) \times (n_1 + n_2 + \ldots + n_s).$$

Ejemplo 1.2 *Sea la matriz*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix}.$$

Obtenga una partición de A de tamaños

$$(i)\ (2+2) \times (2+2+1) \qquad y \qquad (ii)\ (1+2+1) \times (2+3).$$

Solución.

En el primer caso,

$$A = \left[\begin{array}{cc:cc:c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hdashline \ldots & \ldots & \ldots & \ldots & \ldots \\ \hdashline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{array}\right],$$

la cual puede ser escrita de la forma

$$A = \begin{bmatrix} A_{11} & \vdots & A_{12} & \vdots & A_{13} \\ \cdots & . & \cdots & . & \cdots \\ A_{21} & \vdots & A_{22} & \vdots & A_{23} \end{bmatrix},$$

donde

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad A_{12} = \begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix}, \quad A_{13} = \begin{bmatrix} a_{15} \\ a_{25} \end{bmatrix},$$
$$A_{21} = \begin{bmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}, \quad A_{22} = \begin{bmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix}, \quad A_{23} = \begin{bmatrix} a_{35} \\ a_{45} \end{bmatrix}.$$

En el segundo caso,

$$A = \begin{bmatrix} a_{11} & a_{12} & \vdots & a_{13} & a_{14} & a_{15} \\ \cdots & \cdots & . & \cdots & \cdots & \cdots \\ a_{21} & a_{22} & \vdots & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & \vdots & a_{33} & a_{34} & a_{35} \\ \cdots & \cdots & . & \cdots & \cdots & \cdots \\ a_{41} & a_{42} & \vdots & a_{43} & a_{44} & a_{45} \end{bmatrix},$$

la cual puede ser escrita de la forma

$$A = \begin{bmatrix} A'_{11} & \vdots & A'_{12} \\ \cdots & . & \cdots \\ A'_{21} & \vdots & A'_{22} \\ \cdots & . & \cdots \\ A'_{31} & \vdots & A'_{32} \end{bmatrix},$$

donde

$$A'_{11} = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}, \quad A'_{12} = \begin{bmatrix} a_{13} & a_{14} & a_{15} \end{bmatrix},$$
$$A'_{21} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \quad A'_{22} = \begin{bmatrix} a_{23} & a_{24} & a_{25} \\ a_{33} & a_{34} & a_{35} \end{bmatrix},$$
$$A'_{31} = \begin{bmatrix} a_{41} & a_{42} \end{bmatrix}, \quad A'_{32} = \begin{bmatrix} a_{43} & a_{44} & a_{45} \end{bmatrix}.$$

Definición 1.29 Submatriz principal

Si A es una matriz cuadrada de tamaño $n \times n$, se le llama submatriz principal a toda submatriz de A formada eligiendo los mismos índices para las filas y las columnas. Si k es un subconjunto propio de $\{1, 2, \ldots, n\}$, denotamos por $A_{(k)}$ la submatriz principal de A resultante de seleccionar las respectivas filas y columnas indicadas por k.

El hecho de tomar las mismas filas y columnas es equivalente a que los elementos de la diagonal principal de la submatriz han de ser elementos que ya formaban parte de la diagonal principal de la matriz original. Luego si A es simétrica, cualquier submatriz principal también es simétrica.

Nota 1.2 El número total de submatrices principales de orden r que se pueden obtener de una matriz A de orden n es igual a $\binom{n}{r}$.

Ejemplo 1.3 *Obtenga algunas submatrices principales de la matriz*

$$A = \begin{bmatrix} -5 & 3 & 5 & 1 & 0 \\ -4 & 8 & 10 & 0 & -1 \\ 2 & 13 & 11 & 2 & 1 \\ 0 & 1 & 1 & 3 & 2 \\ 3 & 1 & 0 & 7 & 5 \end{bmatrix}.$$

Solución.

Las siguientes matrices son submatrices principales de orden 2 de A:

$$A_{(\{1,2\})} = \begin{bmatrix} -5 & 3 \\ -4 & 8 \end{bmatrix}, \qquad A_{(\{1,3\})} = \begin{bmatrix} -5 & 5 \\ 2 & 11 \end{bmatrix};$$

el número total de submatrices de orden 2 que se pueden obtener de A es $\binom{5}{2} = 10$. Algunas submatrices principales de orden 3 de A son

$$A_{(\{1,4,5\})} = \begin{bmatrix} -5 & 1 & 0 \\ 0 & 3 & 2 \\ 3 & 7 & 5 \end{bmatrix}, \qquad A_{(\{3,4,5\})} = \begin{bmatrix} 11 & 2 & 1 \\ 1 & 3 & 2 \\ 0 & 7 & 5 \end{bmatrix};$$

el número total de submatrices de orden 3 que se pueden obtener de A es $\binom{5}{3} = 10$. Nótese que la submatriz

$$B = \begin{bmatrix} -5 & 1 & 0 \\ -4 & 0 & -1 \\ 2 & 2 & 1 \end{bmatrix}$$

no es principal, porque se ha obtenido con las filas 1, 2 y 3 y con las columnas 1, 4 y 5.

Definición 1.30 Submatriz angular

La submatriz principal que es formada con las primeras k filas y k columnas de la matriz A y que denotaremos por $A_{[k]}$, siendo k el orden de la submatriz, se denomina submatriz angular.

Si A es la matriz cuadrada de tamaño $n \times n$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

entonces las submatrices angulares de A vienen dadas por

$$A_{[1]} = \begin{bmatrix} a_{11} \end{bmatrix},\ A_{[2]} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},\ A_{[3]} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},\ \dots,\ A_{[n]} = A.$$

Así, en el Ejemplo 1.3, la submatriz $A_{(\{1,2\})}$ sería una submatriz angular, concretamente $A_{[2]}$.

Definición 1.31 *Dos matrices A y B están particionadas idénticamente si las submatrices resultantes contienen el mismo número de filas y de columnas y si, además, las partes correspondientes tienen el mismo tamaño. Por tanto, dos matrices particionadas idénticamente son iguales si y solo si las submatrices correspondientes son iguales.*

Definición 1.32 Suma de matrices particionadas

Sean A y B dos matrices particionadas idénticamente. Entonces la suma de A y B tendrá igual partición. En este caso, cada bloque de $A+B$ es obtenido de los correspondientes bloques de A y de B, es decir,

$$A+B=[A_{ij}]+[B_{ij}] = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \dots & A_{1s}+B_{1s} \\ A_{21}+B_{21} & A_{22}+B_{22} & \dots & A_{2s}+B_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1}+B_{r1} & A_{r2}+B_{r2} & \dots & A_{rs}+B_{rs} \end{bmatrix},$$

donde las submatrices A_{ij} y B_{ij} son de tamaño $m_i \times n_j$.

Definición 1.33 Multiplicación por un escalar

Si A es una matriz real de tamaño $m \times n$ particionada y $\alpha \in \mathbb{R}$, entonces la multiplicación de un escalar por A es una matriz real de tamaño $m \times n$ obtenida de multiplicar cada bloque de A por el número α. En otras palabras,

$$\alpha A = \alpha \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{bmatrix} = \begin{bmatrix} \alpha A_{11} & \alpha A_{12} & \dots & \alpha A_{1s} \\ \alpha A_{21} & \alpha A_{22} & \dots & \alpha A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha A_{r1} & \alpha A_{r2} & \dots & \alpha A_{rs} \end{bmatrix},$$

donde A_{ij} son submatrices de tamaño $m_i \times n_j$ para $i = 1, 2, \dots, r$, $j = 1, 2, \dots, s$, donde $m_1 + m_2 + \dots + m_r = m$ y $n_1 + n_2 + \dots + n_s = n$.

Definición 1.34 Transpuesta

Sea A una matriz real de tamaño $m \times n$ particionada de alguna manera. Entonces la transpuesta de A que se escribe A^t es una matriz real

de tamaño $n \times m$ obtenida de intercambiar los renglones por las columnas en cada uno de los bloques A_{ij}. En otras palabras,

$$\text{si } A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{bmatrix}, \text{ entonces } A^t = \begin{bmatrix} A_{11}^t & A_{21}^t & \dots & A_{r1}^t \\ A_{12}^t & A_{22}^t & \dots & A_{r2}^t \\ \vdots & \vdots & \ddots & \vdots \\ A_{1s}^t & A_{2s}^t & \dots & A_{rs}^t \end{bmatrix},$$

donde A_{ij} son submatrices de tamaño $m_i \times n_j$ para $i = 1, 2, \dots, r$, $j = 1, 2, \dots, s$, donde $m_1 + m_2 + \dots + m_r = m$ y $n_1 + n_2 + \dots + n_s = n$.

Ejemplo 1.4 *Obtenga la transpuesta de A para una partición de tamaño $(2+1) \times (2+2)$:*

$$A = \begin{bmatrix} 8 & 9 & 3 & -5 \\ 20 & 10 & -10 & 8 \\ 21 & -5 & 13 & 5 \end{bmatrix}.$$

Solución.

Consideremos la partición $(2+1) \times (2+2)$ de A, es decir,

$$A = \left[\begin{array}{ccc} A_{11} & \vdots & A_{12} \\ \dots & . & \dots \\ A_{21} & \vdots & A_{22} \end{array}\right] = \left[\begin{array}{cccc} 8 & 9 & \vdots & 3 & -5 \\ 20 & 10 & \vdots & -10 & 8 \\ \dots & \dots & . & \dots & \dots \\ 21 & -5 & \vdots & 13 & 5 \end{array}\right].$$

Luego,

$$A_{11}^t = \begin{bmatrix} 8 & 9 \\ 20 & 10 \end{bmatrix}^t = \begin{bmatrix} 8 & 20 \\ 9 & 10 \end{bmatrix}, \quad A_{12}^t = \begin{bmatrix} 3 & -5 \\ -10 & 8 \end{bmatrix}^t = \begin{bmatrix} 3 & -10 \\ -5 & 8 \end{bmatrix},$$

$$A_{21}^t = \begin{bmatrix} 21 & -5 \end{bmatrix}^t = \begin{bmatrix} 21 \\ -5 \end{bmatrix}, \quad A_{22}^t = \begin{bmatrix} 13 & 5 \end{bmatrix}^t = \begin{bmatrix} 13 \\ 5 \end{bmatrix},$$

por consiguiente,

$$\begin{bmatrix} 8 & 9 & \vdots & 3 & -5 \\ 20 & 10 & \vdots & -10 & 8 \\ \cdots & \cdots & \cdot & \cdots & \cdots \\ 21 & -5 & \vdots & 13 & 5 \end{bmatrix}^t = \begin{bmatrix} 8 & 20 & \vdots & 21 \\ 9 & 10 & \vdots & -5 \\ \cdots & \cdots & \cdot & \cdots \\ 3 & -10 & \vdots & 13 \\ -5 & 8 & \vdots & 5 \end{bmatrix}.$$

Teorema 1.19 Multiplicación

Sean A y B matrices particionadas compatibles para el producto, digamos entonces

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{bmatrix} \quad y \quad B = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1t} \\ B_{21} & B_{22} & \dots & B_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ B_{s1} & B_{s2} & \dots & B_{st} \end{bmatrix}.$$

Por tanto, la multiplicación de las dos matrices es

$$A.B = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1t} \\ C_{21} & C_{22} & \dots & C_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ C_{r1} & C_{r2} & \dots & C_{rt} \end{bmatrix},$$

donde $C_{ik} = \sum_{j=1}^{s} A_{ij}B_{jk}$.

Demostración.

Consideremos $(m_1+m_2+\ldots+m_r)\times(n_1+n_2+\ldots+n_s)$ una partición de A y $(n_1+n_2+\ldots+n_s)\times(p_1+p_2+\ldots+p_t)$ una partición de B. Entonces,

$$A_{ij} = \left[a_{hl}\right]_{\substack{m_1+m_2+\ldots+m_{i-1}+1 \le h \le m_1+m_2+\ldots+m_{i-1}+m_i \\ n_1+n_2+\ldots+n_{j-1}+1 \le l \le n_1+n_2+\ldots+n_{j-1}+n_j}}$$

$$B_{jk} = \left[b_{lq}\right]_{\substack{n_1+n_2+\ldots+n_{j-1}+1 \le l \le n_1+n_2+\ldots+n_{j-1}+n_j \\ p_1+p_2+\ldots+p_{k-1}+1 \le q \le p_1+p_2+\ldots+p_{k-1}+p_k}}$$

$$A_{ij}B_{jk} = \left[\sum_{l=n_1+\ldots+n_{j-1}+1}^{n_1+\ldots+n_{j-1}+n_j} a_{hl}b_{lq}\right]_{\substack{m_1+\ldots+m_{i-1}+1 \le h \le m_1+\ldots+m_{i-1}+m_i \\ p_1+p_2+\ldots+p_{k-1}+1 \le q \le p_1+p_2+\ldots+p_{k-1}+p_k}},$$

y por tanto

$$\begin{aligned} C_{ik} &= \sum_{j=1}^{s} A_{ij}B_{jk} \\ &= \left[\sum_{l=1}^{n} a_{hl}b_{lq}\right]_{\substack{m_1+m_2+\ldots+m_{i-1}+1 \le h \le m_1+m_2+\ldots+m_{i-1}+m_i \\ p_1+p_2+\ldots+p_{k-1}+1 \le q \le p_1+p_2+\ldots+p_{k-1}+p_k}}, \end{aligned}$$

es decir, C_{ik} es el bloque (i,k) correspondiente a la partición

$$(m_1+m_2+\ldots+m_r)\times(p_1+p_2+\ldots+p_t)$$

de la matriz $A.B$. ■

Ejemplo 1.5 *Calcular $A.B$ asumiendo una partición para A de tamaño $(2+1)\times(2+2)$ y otra para B de tamaño $(2+2)\times(2+2)$, donde*

$$A = \begin{bmatrix} 8 & 9 & 3 & -5 \\ 20 & 10 & -10 & 8 \\ 21 & -5 & 13 & 5 \end{bmatrix} \quad y \quad B = \begin{bmatrix} 2 & 1 & 3 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 2 & -1 \end{bmatrix}.$$

Solución.

La partición de tamaño $(2+1)\times(2+2)$ para A es

$$A = \begin{bmatrix} A_{11} & \vdots & A_{12} \\ \cdots & . & \cdots \\ A_{21} & \vdots & A_{22} \end{bmatrix} = \left[\begin{array}{cc:cc} 8 & 9 & 3 & -5 \\ 20 & 10 & -10 & 8 \\ \hdashline \cdots & \cdots & \cdots & \cdots \\ 21 & -5 & 13 & 5 \end{array}\right]$$

y la partición de tamaño $(2+2)\times(2+2)$ para B es

$$B = \begin{bmatrix} B_{11} & \vdots & B_{12} \\ \cdots & . & \cdots \\ B_{21} & \vdots & B_{22} \end{bmatrix} = \left[\begin{array}{cc:cc} 2 & 1 & 3 & 1 \\ 1 & 1 & -1 & -1 \\ \hdashline 1 & -1 & 1 & 1 \\ -1 & 1 & 2 & -1 \end{array}\right]. \tag{1.11}$$

Luego,

$$\begin{aligned} A_{11}B_{11}+A_{12}B_{21} &= \begin{bmatrix} 25 & 17 \\ 50 & 30 \end{bmatrix} + \begin{bmatrix} 8 & -8 \\ -18 & 18 \end{bmatrix} = \begin{bmatrix} 33 & 9 \\ 32 & 48 \end{bmatrix}, \\ A_{11}B_{12}+A_{12}B_{22} &= \begin{bmatrix} 15 & -1 \\ 50 & 10 \end{bmatrix} + \begin{bmatrix} -7 & 8 \\ 6 & -18 \end{bmatrix} = \begin{bmatrix} 8 & 7 \\ 56 & -8 \end{bmatrix}, \\ A_{21}B_{11}+A_{22}B_{21} &= \begin{bmatrix} 37 & 16 \end{bmatrix} + \begin{bmatrix} 8 & -8 \end{bmatrix} = \begin{bmatrix} 45 & 8 \end{bmatrix}, \\ A_{21}B_{12}+A_{22}B_{22} &= \begin{bmatrix} 68 & 26 \end{bmatrix} + \begin{bmatrix} 23 & 8 \end{bmatrix} = \begin{bmatrix} 91 & 34 \end{bmatrix}. \end{aligned}$$

Por consiguiente,

$$A.B = \left[\begin{array}{cc:cc} 33 & 9 & 8 & 7 \\ 32 & 48 & 56 & -8 \\ \hdashline \cdots & \cdots & \cdots & \cdots \\ 45 & 8 & 91 & 34 \end{array}\right].$$

1.5.2 Determinantes de matrices particionadas

En esta sección se muestran algunos resultados para encontrar el determinante de una matriz cuadrada particionada en bloques.

Teorema 1.20 *Sea A una matriz real de tamaño $n \times n$ particionada como sigue:*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{1.12}$$

donde A_{ij} son submatrices de tamaño $n_i \times n_j$ para $i,\ j = 1, 2$, con $n_1 + n_2 = n$, $0 < n_1 < n$. Si $A_{12} = \mathrm{O}$ o si $A_{21} = \mathrm{O}$, entonces

$$\det(A) = \det(A_{11}) \det(A_{22}).$$

Demostración.

El resultado se prueba por inducción sobre k. Sin pérdida de generalidad, supongamos que $A_{12} = \mathrm{O}$ y asumamos que el teorema es válido para todas las matrices de tamaño $(n-1) \times (n-1)$ de la forma apropiada. Como el determinante de una matriz A de tamaño $n \times n$, se puede calcular mediante la expansión de Laplace (método de cofactores) por la primera fila, como sigue

$$\begin{aligned} \det A &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} M_{1j}(A) \qquad (1.13) \\ &= a_{11} M_{11}(A) - a_{12} M_{12}(A) + \ldots + (-1)^{n+1} a_{1n} M_{1n}(A), \end{aligned}$$

donde cada uno de los menores complementarios $M_{1j}(A)$ son de la forma

$$M_{1j}(A) = \begin{vmatrix} \left[A_{11}^{[j]}\right] & \mathrm{O} \\ [A_{21}]_j & A_{22} \end{vmatrix}, \qquad j = 1, 2, \ldots, k.$$

Aquí, $\left[A_{11}^{[j]}\right]$ se consigue borrando de A_{11} la primera fila y la j-ésima columna y $[A_{21}]_j$ se obtiene de suprimir de A_{21} la j-ésima columna. Por inducción sobre k,

$$M_{1j}(A) = \det\left(\left[A_{11}^{[j]}\right]\right) \det\left(A_{22}\right).$$

Si se reemplaza en (1.13), se tiene que

$$\begin{aligned} \det A &= \left[\sum_{j=1}^{k} (-1)^{1+j} a_{1j} \det\left(\left[A_{11}^{[j]}\right]\right)\right] \det(A_{22}) \\ &= \det(A_{11}) \det(A_{22}), \end{aligned}$$

como se deseaba. ■

Definición 1.35 Complemento de Schur

Sea A una matriz de tamaño $n \times n$ particionada como en (1.12)*, si A_{11} es una submatriz principal no singular de A, se define el complemento de Schur de A_{11} en A, denotado $\big(A/A_{11}\big)$ como sigue:*

$$\big(A/A_{11}\big) = A_{22} - A_{21}A_{11}^{-1}A_{12}. \tag{1.14}$$

Teorema 1.21 Fórmula de Schur

Sea A una matriz real de tamaño $n \times n$ particionada como en (1.12)*. Si A_{11} y A_{22} son submatrices cuadradas, entonces:*

1. *Si A_{11} es no singular se verifica que*

$$\det A = \det(A_{11}) \det\big[\big(A/A_{11}\big)\big], \tag{1.15}$$

   donde la submatriz $\big(A/A_{11}\big) = A_{22} - A_{21}A_{11}^{-1}A_{12}$.

2. *Si A_{22} es no singular se cumple que*

$$\det A = \det(A_{22}) \det\big[\big(A/A_{22}\big)\big], \tag{1.16}$$

   donde la submatriz $\big(A/A_{22}\big) = A_{11} - A_{12}A_{22}^{-1}A_{21}$.

Demostración.

1. Consideremos que en la partición (1.12) la submatriz A_{11} es no singular, luego la matriz A se puede factorizar como sigue

$$\underbrace{\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} A_{11} & \mathrm{O} \\ A_{21} & \big(A/A_{11}\big) \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} I & A_{11}^{-1}A_{12} \\ \mathrm{O} & I \end{bmatrix}}_{U}.$$

   Fácilmente, el lector puede probar que $A = LU$. Por otra parte, como $\det(LU) = \det(L)\det(U)$ para matrices cuadradas L y U, por el Teorema 1.20 se tiene que

$$\det A = \det(LU) = \det(A_{11}) \det\big[\big(A/A_{11}\big)\big].$$

2. Queda como ejercicio para el lector. ■

Ejemplo 1.6 *Obtenga el determinante de la matriz B dada en el Ejemplo 1.5, asumiendo una partición de tamaño* $(2+2)\times(2+2)$.

Solución.

Usando la partición dada en (1.11) de la matriz B, se tiene que B_{11} es no singular ya que su determinante es 1. Luego

$$B_{21}B_{11}^{-1}B_{12} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 9 & 5 \\ -9 & -5 \end{bmatrix}.$$

Nótese que esta matriz no es invertible, por otra parte,

$$\left(B/B_{11}\right) = B_{22} - B_{21}B_{11}^{-1}B_{12} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} - \begin{bmatrix} 9 & 5 \\ -9 & -5 \end{bmatrix} = \begin{bmatrix} -8 & -4 \\ 11 & 4 \end{bmatrix},$$

tiene determinante igual a 12. Usando la expresión (1.15) se tiene que

$$\det B = 1 \cdot 12 = 12.$$

1.5.3 Inversas de matrices particionadas

Para determinar la inversa de una matriz cuadrada, usualmente se emplea el método de Gauss-Jordan o el método del determinante y la matriz adjunta. En esta sección se ilustra la manera en que se pueden calcular las inversas de las matrices usando particiones.

Teorema 1.22 *Sea A una matriz real no singular particionada como en* (1.12)*, con A_{11} y A_{22} submatrices cuadradas no singulares. Entonces:*

1. *Si $A_{12} = \mathrm{O}$, la inversa de A es*

$$\begin{bmatrix} A_{11} & \mathrm{O} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & \mathrm{O} \\ -A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix}.$$

2. *Si $A_{21} = \mathrm{O}$, la inversa de la matriz dada en* (1.12) *es*

$$\begin{bmatrix} A_{11} & A_{12} \\ \mathrm{O} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ \mathrm{O} & A_{22}^{-1} \end{bmatrix}.$$

Demostración.
Si A^{-1} es la inversa de A, entonces se debe cumplir que $AA^{-1} = I$.

1. Supóngase que A_{12} es una submatriz nula, entonces

$$\begin{aligned} AA^{-1} &= \begin{bmatrix} A_{11} & \mathrm{O} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & \mathrm{O} \\ -A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}A_{11}^{-1} & \mathrm{O} \\ A_{21}A_{11}^{-1} - A_{22}A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}A_{22}^{-1} \end{bmatrix} = \begin{bmatrix} I_{n_1} & \mathrm{O} \\ \mathrm{O} & I_{n_2} \end{bmatrix}. \end{aligned}$$

2. De manera análoga, cuando $A_{21} = \mathrm{O}$, se tiene que

$$\begin{aligned} AA^{-1} &= \begin{bmatrix} A_{11} & A_{12} \\ \mathrm{O} & A_{22} \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ \mathrm{O} & A_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}A_{11}^{-1} & A_{12}A_{22}^{-1} - A_{11}A_{11}^{-1}A_{12}A_{22}^{-1} \\ \mathrm{O} & A_{22}A_{22}^{-1} \end{bmatrix} = \begin{bmatrix} I_{n_1} & \mathrm{O} \\ \mathrm{O} & I_{n_2} \end{bmatrix}. \end{aligned}$$

Nótese que en ambos casos se obtiene la matriz identidad. ■

Teorema 1.23 Inversa de una matriz particionada

Sea A una matriz no singular particionada como sigue:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

donde A_{ij} son submatrices de tamaño $n_i \times n_j$ para $i,\ j = 1, 2$, con $n_1 + n_2 = n$ y $0 < n_1 < n$, denotando A^{-1} por G y particionando G como

$$G = \begin{bmatrix} \big(A/A_{22}\big)^{-1} & G_{12} \\ G_{21} & \big(A/A_{11}\big)^{-1} \end{bmatrix},$$

donde G_{ij} son submatrices de tamaño $n_i \times n_j$ para $i,\ j = 1, 2$. Si $\det\big(A_{11}\big) \neq 0$ y $\det\big(A_{22}\big) \neq 0$, se tienen los siguientes resultados:

1. $\big(A/A_{22}\big) = A_{11} - A_{12}A_{22}^{-1}A_{21}$ *existe y es no singular.*

2. $\big(A/A_{11}\big) = A_{22} - A_{21}A_{11}^{-1}A_{12}$ *existe y es no singular.*

3. A^{-1} *puede escribirse como*

$$a)\ A^{-1} = \begin{bmatrix} \big(A/A_{22}\big)^{-1} & -\big(A/A_{22}\big)^{-1}A_{12}A_{22}^{-1} \\ -\big(A/A_{11}\big)^{-1}A_{21}A_{11}^{-1} & \big(A/A_{11}\big)^{-1} \end{bmatrix},$$

$$b)\ A^{-1} = \begin{bmatrix} \big(A/A_{22}\big)^{-1} & -A_{11}^{-1}A_{12}\big(A/A_{11}\big)^{-1} \\ -A_{22}^{-1}A_{21}\big(A/A_{22}\big)^{-1} & \big(A/A_{11}\big)^{-1} \end{bmatrix}.$$

Demostración.

Para probar que $\big(A/A_{22}\big)$ es no singular, se multiplica A por la izquierda por la matriz no singular A_1^*

$$A_1^* = \begin{bmatrix} I_{n_1} & -A_{12}A_{22}^{-1} \\ \mathrm{O} & A_{22}^{-1} \end{bmatrix}.$$

Fácilmente, el lector puede probar que $A_1^*A = C$, donde

$$C = \begin{bmatrix} \big(A/A_{22}\big) & \mathrm{O} \\ A_{22}^{-1}A_{21} & I_{n_2} \end{bmatrix},$$

pero por el Teorema 1.20 se tiene que

$$\det(C) = \det\big[\big(A/A_{22}\big)\big]\det\big(I_{n_2}\big) = \det(A_1^*A) = \det(A_1^*)\det(A) \neq 0.$$

En consecuencia,

$$\big(A/A_{22}\big) = A_{11} - A_{12}A_{22}^{-1}A_{21}$$

es una matriz no singular.

Para probar que $\big(A/A_{11}\big)$ es no singular, se multiplica la matriz A por la izquierda por la matriz no singular:

$$A_2^* = \begin{bmatrix} A_{11}^{-1} & \mathrm{O} \\ -A_{21}A_{11}^{-1} & I_{n_2} \end{bmatrix}.$$

Por otra parte, del resultado $GA = I_n$ se tiene que

$$\begin{bmatrix} \big(A/A_{22}\big)^{-1} & G_{12} \\ G_{21} & \big(A/A_{11}\big)^{-1} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_{n_1} & \mathrm{O} \\ \mathrm{O} & I_{n_2} \end{bmatrix}$$

y de estas se obtienen las siguientes cuatro ecuaciones matriciales

$$\begin{aligned}
\big(A/A_{22}\big)^{-1}A_{11} + G_{12}A_{21} &= I_{n_1}, && \text{(1.17a)}\\
\big(A/A_{22}\big)^{-1}A_{12} + G_{12}A_{22} &= \mathrm{O}, && \text{(1.17b)}\\
G_{21}A_{11} + \big(A/A_{11}\big)^{-1}A_{21} &= \mathrm{O}, && \text{(1.17c)}\\
G_{21}A_{12} + \big(A/A_{11}\big)^{-1}A_{22} &= I_{n_2}. && \text{(1.17d)}
\end{aligned}$$

Si se multiplica por la derecha de la ecuación (1.17b) por A_{22}^{-1}, se tiene

$$G_{12} = -\big(A/A_{22}\big)^{-1}A_{12}A_{22}^{-1}.$$

Si se reemplaza en la ecuación (1.17a) y se factoriza, se obtiene

$$\big(A/A_{22}\big)^{-1}\big[A_{11} - A_{12}A_{22}^{-1}A_{21}\big] = I_{n_1}. \tag{1.18}$$

Así, $\big(A/A_{22}\big)$ existe y es igual a $A_{11} - A_{12}A_{22}^{-1}A_{21}$.

De manera análoga, si se utilizan las ecuaciones (1.17c) y (1.17d), se puede probar que

$$\big(A/A_{11}\big)^{-1}\big[A_{22} - A_{21}A_{11}^{-1}A_{12}\big] = I_{n_2}, \tag{1.19}$$

es decir, $\big(A/A_{11}\big)$ existe y es igual a $A_{22} - A_{21}A_{11}^{-1}A_{12}$.

De reemplazar en G_{12} y G_{21}, se sigue la prueba de 3. *a*). ■

Corolario 1.23.1 *Sean A y G como en el Teorema 1.23, entonces se satisfacen las siguientes condiciones:*

1. $\big(A/A_{22}\big)^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}\big(A/A_{11}\big)^{-1}A_{21}A_{11}^{-1}$,
2. $\big(A/A_{11}\big)^{-1} = A_{22}^{-1} + A_{22}^{-1}A_{21}\big(A/A_{22}\big)^{-1}A_{12}A_{22}^{-1}$,
3. $A_{22}^{-1}A_{21}\big(A/A_{22}\big)^{-1} = \big(A/A_{11}\big)^{-1}A_{21}A_{11}^{-1}$,
4. $A_{11}^{-1}A_{12}\big(A/A_{11}\big)^{-1} = \big(A/A_{22}\big)^{-1}A_{12}A_{22}^{-1}$,

donde $\big(A/A_{22}\big) = A_{11} - A_{12}A_{22}^{-1}A_{21}$ *y* $\big(A/A_{11}\big) = A_{22} - A_{21}A_{11}^{-1}A_{12}$.

Demostración.
Como A_{11} es invertible, la matriz A se puede factorizar como sigue

$$\underbrace{\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} A_{11} & \mathrm{O} \\ A_{21} & (A/A_{11}) \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} I_{n_1} & A_{11}^{-1}A_{12} \\ \mathrm{O} & I_{n_2} \end{bmatrix}}_{U}.$$

Luego $A^{-1} = G = U^{-1}L^{-1}$, y por el Teorema 1.22, se obtiene

$$\begin{aligned} A^{-1} &= \begin{bmatrix} I_{n_1} & -A_{11}^{-1}A_{12} \\ \mathrm{O} & I_{n_2} \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & \mathrm{O} \\ -(A/A_{11})^{-1}A_{21}A_{11}^{-1} & (A/A_{11})^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}(A/A_{11})^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}(A/A_{11})^{-1} \\ -(A/A_{11})^{-1}A_{21}A_{11}^{-1} & (A/A_{11})^{-1} \end{bmatrix}. \end{aligned}$$

Como en el Teorema 1.23 se obtuvo A^{-1}, comparando los términos se obtienen las condiciones 1. y 3.

$$\begin{aligned} (A/A_{22})^{-1} &= A_{11}^{-1} + A_{11}^{-1}A_{12}(A/A_{11})^{-1}A_{21}A_{11}^{-1}, \\ A_{22}^{-1}A_{21}(A/A_{22})^{-1} &= (A/A_{11})^{-1}A_{21}A_{11}^{-1}. \end{aligned}$$

Por otra parte, como A_{22} es no singular, la matriz A puede factorizarse de la siguiente manera

$$\underbrace{\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} (A/A_{22}) & A_{12} \\ \mathrm{O} & A_{22} \end{bmatrix}}_{R} \underbrace{\begin{bmatrix} I_{n_1} & \mathrm{O} \\ A_{22}^{-1}A_{21} & I_{n_2} \end{bmatrix}}_{S}.$$

Por lo tanto, $A^{-1} = G = S^{-1}R^{-1}$, y en virtud del Teorema 1.22, se tiene que

$$\begin{aligned} A^{-1} &= \begin{bmatrix} I_{n_1} & \mathrm{O} \\ -A_{22}^{-1}A_{21} & I_{n_2} \end{bmatrix} \begin{bmatrix} (A/A_{22})^{-1} & -(A/A_{22})^{-1}A_{12}A_{22}^{-1} \\ \mathrm{O} & A_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} (A/A_{22})^{-1} & -(A/A_{22})^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}(A/A_{22})^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}(A/A_{22})^{-1}A_{12}A_{22}^{-1} \end{bmatrix}. \end{aligned}$$

Aquí se comparan de nuevo los términos con los de la matriz A^{-1}, para obtener las condiciones 2. y 4.

$$\begin{aligned} (A/A_{11})^{-1} &= A_{22}^{-1} + A_{22}^{-1}A_{21}(A/A_{22})^{-1}A_{12}A_{22}^{-1}, \\ A_{11}^{-1}A_{12}(A/A_{11})^{-1} &= (A/A_{22})^{-1}A_{12}A_{22}^{-1} \end{aligned}$$

y el corolario queda probado. ■

Ejemplo 1.7

Obtenga la inversa de la matriz B dada en el Ejemplo 1.5, asumiendo una partición de tamaño $(2+2)\times(2+2)$.

Solución.

En el Ejemplo 1.6, se obtuvo que $\det B = 12$, es decir B es invertible. Asumiendo la partición dada en (1.11) de la matriz B, se tiene que

$$B_{11}^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}, \qquad B_{22}^{-1} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^{-1} = \frac{1}{3}\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix},$$

$$B_{12}B_{22}^{-1}B_{21} = \begin{bmatrix} 3 & 1 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Aunque $B_{12}B_{22}^{-1}B_{21}$ no es invertible, el primer bloque $\big(B/B_{22}\big)^{-1}$ de la matriz B^{-1} viene dado por

$$\big(B/B_{22}\big)^{-1} = \big[B_{11} - B_{12}B_{22}^{-1}B_{21}\big]^{-1} = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}^{-1} = \frac{1}{4}\begin{bmatrix} 0 & 2 \\ 2 & -1 \end{bmatrix}.$$

La submatriz $G_{21} = -B_{22}^{-1}B_{21}\big(B/B_{22}\big)^{-1}$ está dada por

$$G_{21} = -\frac{1}{4}\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 0 & 2 \\ 2 & -1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 0 & 0 \\ 2 & -3 \end{bmatrix} \qquad \text{y}$$

$$B_{21}B_{11}^{-1}B_{12} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 3 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 9 & 5 \\ -9 & -5 \end{bmatrix}.$$

A pesar de que $B_{21}B_{11}^{-1}B_{12}$ no es invertible, el bloque $\big(B/B_{11}\big)^{-1}$ se obtiene como sigue

$$\big(B/B_{11}\big)^{-1} = \big[B_{22} - B_{21}B_{11}^{-1}B_{12}\big]^{-1} = \begin{bmatrix} -8 & -4 \\ 11 & 4 \end{bmatrix}^{-1} = \frac{1}{12}\begin{bmatrix} 4 & 4 \\ -11 & -8 \end{bmatrix},$$

y finalmente, se tiene que $G_{12} = -B_{11}^{-1}B_{12}\big(B/B_{11}\big)^{-1}$, es decir,

$$G_{12} = -\frac{1}{12}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 3 & 1 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} 4 & 4 \\ -11 & -8 \end{bmatrix} = \frac{1}{12}\begin{bmatrix} 6 & 0 \\ -13 & -4 \end{bmatrix}.$$

Por tanto,

$$\begin{bmatrix} 2 & 1 & 3 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 2 & -1 \end{bmatrix}^{-1} = \frac{1}{12}\begin{bmatrix} 0 & 6 & 6 & 0 \\ 6 & -3 & -13 & -4 \\ 0 & 0 & 4 & 4 \\ 6 & -9 & -11 & -8 \end{bmatrix}.$$

Ejercicios 1.1

1. Utilizando particiones, encuentre el determinante y la inversa de la matriz dada en el Ejemplo (1.3).

2. Si A y C son no singulares, pruebe que

   i) $\det(I + AB) = \det(I + BA)$.

   ii) $\det(A + CBC^t) = \det(A)\det(I + BC^tA^{-1}C)$.

3. Demuestre que

$$\begin{bmatrix} P & Q \\ R & S \end{bmatrix} \begin{bmatrix} I & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix} \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} = \begin{bmatrix} P \\ R \end{bmatrix} \begin{bmatrix} X & Y \end{bmatrix}.$$

4. Muestre que la inversa de una matriz no singular particionada es

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix} E^{-1} \begin{bmatrix} -CA^{-1} & I \end{bmatrix},$$

   donde A es no singular y $E = D - CA^{-1}B$.

1.6 Espacio vectorial

Los conjuntos $\mathbb{R}^2$ (vectores en el plano) y $\mathbb{R}^3$ (vectores en el espacio) junto con las operaciones de suma de vectores y multiplicación por un escalar se llaman *espacios vectoriales*. Las propiedades algebraicas de un espacio vectorial arbitrario son muy semejantes a las de los elementos de $\mathbb{R}^2$ y $\mathbb{R}^3$. En consecuencia, se acostumbra llamar *vectores* también a los elementos de un espacio vectorial arbitrario.

Definición 1.36 *Un espacio vectorial real*[2] $\mathbb{V}$ *es un conjunto no vacío de vectores, dotado de dos operaciones:*

Suma	**Multiplicación por un escalar**
$\mathbb{V} \times \mathbb{V} \to \mathbb{V}$	$\mathbb{R} \times \mathbb{V} \to \mathbb{V}$
$(\vec{x}, \vec{y}) \to \vec{x} + \vec{y}$	$(\alpha, \vec{x}) \to \alpha\vec{x}$

que satisfacen los diez axiomas enumerados a continuación.

1.6.1 Axiomas de un espacio vectorial

Dado $\mathbb{V}$ un espacio vectorial real, se verifica que

- Para la suma en $\mathbb{V}$:
  - (i) *Clausurativa:* si $\vec{x}, \vec{y} \in \mathbb{V}$ entonces $\vec{x} + \vec{y} \in \mathbb{V}$.
  - (ii) *Asociativa:* para todo $\vec{x}, \vec{y}$ y $\vec{z} \in \mathbb{V}$, $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$.
  - (iii) *Conmutativa:* si $\vec{x}, \vec{y} \in \mathbb{V}$, entonces $\vec{x} + \vec{y} = \vec{y} + \vec{x}$.
  - (iv) *Existencia de elemento neutro:* existe un vector de $\mathbb{V}$ denotado por $\vec{0}$ tal que para todo $\vec{x} \in \mathbb{V}$, $\vec{x} + \vec{0} = \vec{0} + \vec{x} = \vec{x}$.
  - (v) *Elemento opuesto:* si $\vec{x} \in \mathbb{V}$, existe un vector $-\vec{x}$ en $\mathbb{V}$ tal que $\vec{x} + (-\vec{x}) = \vec{0}$.
- Para el producto por un escalar de $\mathbb{R}$:
  - (vi) *Clausurativa:* si $\vec{x} \in \mathbb{V}$ y α es un escalar, entonces $\alpha\vec{x} \in \mathbb{V}$.
  - (vii) *Distributiva respecto a la suma de vectores:* si $\vec{x}, \vec{y} \in \mathbb{V}$ y α es un escalar, entonces $\alpha(\vec{x} + \vec{y}) = \alpha\vec{x} + \alpha\vec{y}$.
  - $(viii)$ *Distributiva respecto a la suma de escalares:* si $\vec{x} \in \mathbb{V}$ y α y β son escalares, entonces $(\alpha + \beta)\vec{x} = \alpha\vec{x} + \beta\vec{x}$.
  - (ix) *Asociativa respecto a la multiplicación de escalares:* si $\vec{x} \in \mathbb{V}$ y α y β son escalares, entonces $\alpha(\beta\vec{x}) = (\alpha\beta)\vec{x}$.
  - (x) *Existencia del elemento unidad:* para cada vector $\vec{x} \in \mathbb{V}$, $1\vec{x} = \vec{x}$.

[2] La palabra "real" significa que los escalares que se usan son números reales.

Definición 1.37 *Sea $\mathbb{V}$ un espacio vectorial y $\mathbb{W}$ un subconjunto no vacío de $\mathbb{V}$. Se dice que $\mathbb{W}$ es un subespacio vectorial de $\mathbb{V}$ si $\mathbb{W}$ dotado de las mismas operaciones definidas en $\mathbb{V}$ es, a su vez, espacio vectorial.*

Teorema 1.24 *Un subconjunto no vacío $\mathbb{W}$ de un espacio vectorial $\mathbb{V}$ es un subespacio vectorial de $\mathbb{V}$ si cumple que:*

i) La suma de elementos de $\mathbb{W}$ es un elemento de $\mathbb{W}$.

ii) El producto de un escalar por un elemento de $\mathbb{W}$ pertenece a $\mathbb{W}$.

Una condición equivalente para que $\mathbb{W}$ sea subespacio vectorial es que para todo par de elementos $\vec{v}$ y $\vec{w}$ de $\mathbb{W}$, y cualesquiera α y β de $\mathbb{R}$, se verifique que $\alpha\vec{v} + \beta\vec{w}$ pertenece a $\mathbb{W}$.

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.38 *Si $\mathbb{U}$ y $\mathbb{W}$ son subespacios de un espacio vectorial real $\mathbb{V}$, entonces se define la suma $\mathbb{U} + \mathbb{W}$ como*

$$\mathbb{U} + \mathbb{W} = \{\vec{u} + \vec{w} \mid \vec{u} \in \mathbb{U}, \vec{w} \in \mathbb{W}\}.$$

Teorema 1.25 *Si $\mathbb{U}$ y $\mathbb{W}$ son subespacios de un espacio vectorial real $\mathbb{V}$, entonces la suma $\mathbb{U} + \mathbb{W}$ es un subespacio de $\mathbb{V}$.*

Demostración.
Se debe probar que $\mathbb{U} + \mathbb{W}$ satisface las condiciones del Teorema 1.24:

i) Si $\vec{u}_1, \vec{u}_2 \in \mathbb{U}$ y $\vec{w}_1, \vec{w}_2 \in \mathbb{W}$, entonces

$$\left(\vec{u}_1 + \vec{w}_1\right) + \left(\vec{u}_2 + \vec{w}_2\right) = \vec{u}_1 + \vec{u}_2 + \vec{w}_1 + \vec{w}_2 \in \mathbb{U} + \mathbb{W}.$$

ii) Si $\alpha \in \mathbb{R}$, entonces

$$\alpha\big(\vec{u}_1 + \vec{w}_1\big) = \alpha\vec{u}_1 + \alpha\vec{w}_1 \in \mathbb{U} + \mathbb{W}.$$

Finalmente, $\vec{0} + \vec{0} \in \mathbb{U} + \mathbb{W}$. Esto prueba que $\mathbb{U} + \mathbb{W}$ es un subespacio. ■

Definición 1.39 *Se dice que $\mathbb{V}$ es una suma directa de $\mathbb{U}$ y $\mathbb{W}$, si todo $\vec{v} \in \mathbb{V}$ tiene una representación única de la forma*

$$\vec{v} = \vec{u} + \vec{w},$$

con $\vec{u} \in \mathbb{U}$ y $\vec{v} \in \mathbb{V}$. Esta suma directa se denotará como $\mathbb{V} = \mathbb{U} \bigoplus \mathbb{W}$.

Teorema 1.26 *Si $\mathbb{U}$ y $\mathbb{W}$ son subespacios no nulos de un espacio vectorial real $\mathbb{V}$, su suma $\mathbb{U}+\mathbb{W}$ es una suma directa si y solo si $\mathbb{U} \bigcap \mathbb{W} = \{\vec{0}\}$.*

Demostración.

Queda como ejercicio para el lector. ■

Definición 1.40 Combinación lineal

Sean $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ vectores en un espacio vectorial real $\mathbb{V}$. Un vector $\vec{v}$ en $\mathbb{V}$ es una combinación lineal de $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ si

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n$$

para ciertos números reales $c_1, c_2, \ldots, c_n$.

Definición 1.41 *Sea $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ un conjunto de vectores en un espacio vectorial $\mathbb{V}$, entonces se dice que S es:*

1. *Linealmente dependiente o ligado si y solo si existen escalares $c_i \in \mathbb{R}$ no todos nulos, tales que*

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n = \vec{0}.$$

2. Linealmente independiente o libre si y solo si no es ligado. Esto es

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n = \vec{0}.$$

Se cumple solo para $c_1 = c_2 = \ldots = c_n = 0$.

Definición 1.42 Espacio generado

Si $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ *es un conjunto de vectores en un espacio vectorial* $\mathbb{V}$*, entonces el conjunto de todos los vectores en* $\mathbb{V}$ *que son combinaciones lineales de los vectores en* S *se denomina espacio generado y se denota por* gen S

$$\text{gen}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} = \{\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_n\vec{v}_n | \alpha_i \in \mathbb{R}\}. \qquad (1.20)$$

1.6.2 Bases

En esta sección se continúa con el estudio de la estructura de un espacio vectorial $\mathbb{V}$ determinando un conjunto mínimo de vectores de $\mathbb{V}$ que describa completamente a $\mathbb{V}$.

Definición 1.43 Base

Si $\mathbb{V}$ *es cualquier espacio vectorial y* $\mathfrak{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ *es un conjunto finito de vectores en* $\mathbb{V}$*, entonces* $\mathfrak{B}$ *se denomina base para* $\mathbb{V}$ *si es un conjunto generador para* $\mathbb{V}$ *con el número más pequeño de elementos en un conjunto generador para* $\mathbb{V}$.

El teorema principal acerca de las bases es:

Teorema 1.27 Base

Sea $\mathfrak{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ *un conjunto de vectores en un espacio vectorial* $\mathbb{V}$*. El conjunto* $\mathfrak{B}$ *es una base para* $\mathbb{V}$ *si y solo si* $\mathfrak{B}$ *es linealmente independiente y genera a* $\mathbb{V}$.

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.44 Dimensión

Si un espacio vectorial $\mathbb{V}$ *tiene una base* $\mathfrak{B}$ *con* n *elementos* ($n \in \mathbb{N}$), *entonces se define a* n *como la dimensión del espacio vectorial* $\mathbb{V}$ *y se escribe*

$$n = \dim \mathbb{V}.$$

Si $\mathbb{V} = \left\{\vec{0}\right\}$, *entonces se tiene que* $\dim \mathbb{V} = 0$.

Teorema 1.28 *Sea* $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ *un conjunto de vectores en un espacio vectorial* $\mathbb{V}$ *de dimensión* n. *Sea*

$$A = \left[\vec{v}_1 \quad \vec{v}_2 \quad \ldots \quad \vec{v}_n\right],$$

entonces S *es un conjunto de vectores linealmente independiente si y solo si* $\det A \neq 0$.

Demostración.
Queda como ejercicio para el lector. ■

Teorema 1.29 *Suponga que* $\dim \mathbb{V} = n$. *Si* $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ *es un conjunto de* m *vectores linealmente independientes en* $\mathbb{V}$, *entonces* $m \leq n$.

Demostración.
Queda como ejercicio para el lector. ■

Teorema 1.30 *Cualesquiera* n *vectores linealmente independientes en un espacio vectorial* $\mathbb{V}$ *de dimensión* n *constituyen una base para* $\mathbb{V}$.

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.45 *Sea* $\mathfrak{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ *una base para el espacio vectorial* $\mathbb{V}$ *de dimensión* n*. Las coordenadas de cualquier vector* $\vec{x} \in \mathbb{V}$ *en* $\mathfrak{B}$ *se relacionan por*

$$\left(\vec{x}\right)_{\mathfrak{B}} = M_{\mathfrak{B}}^{-1}\vec{x}, \tag{1.21}$$

donde $M_{\mathfrak{B}} = [\vec{v}_1 \ \vec{v}_2 \ \ldots \ \vec{v}_n]$.

Ejemplo 1.8 *Muestre que los vectores* ${\vec{v}_1}^t = (2, -1)$ *y* ${\vec{v}_2}^t = (1, 5)$ *forman una base de* $\mathbb{R}^2$*, y luego halle las componentes del vector* $\vec{x}^{\ t} = (7, 4)$ *con relación a esta base.*

Solución.

Fórmese $A = [\vec{v}_1 \ \vec{v}_2]$ y calcúlese su determinante

$$\det A = \left|\vec{v}_1 \quad \vec{v}_2\right| = \begin{vmatrix} 2 & 1 \\ -1 & 5 \end{vmatrix} = 11 \neq 0.$$

Luego, $S = \{\vec{v}_1, \vec{v}_2\}$ es un conjunto de vectores linealmente independientes y como $\mathbb{R}^2$ tiene dimensión dos, se deduce que forman una base.

Para hallar las componentes de $\vec{x}$ en términos de esta base, se hace

$$\begin{aligned} \left(\vec{x}\right)_{\mathfrak{B}} &= M_{\mathfrak{B}}^{-1}\vec{x} = A^{-1}\vec{x} \\ &= \begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 7 \\ 4 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 5 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 7 \\ 4 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 31 \\ 15 \end{bmatrix}. \end{aligned}$$

Estas son las componentes de $\vec{x}$ relativas a la base $\vec{v}_1, \vec{v}_2$.

Teorema 1.31 Cambio de Base

Sean $\mathfrak{B}_1 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ *y* $\mathfrak{B}_2 = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ *bases para el espacio vectorial* $\mathbb{V}$ *de dimensión* n*. Dado cualquier vector* $\vec{x} \in \mathbb{V}$*, sus coordenadas en* $\mathfrak{B}_1$*,* $\left(\vec{x}\right)_{\mathfrak{B}_1}$ *y sus coordenadas en* $\mathfrak{B}_2$*,* $\left(\vec{x}\right)_{\mathfrak{B}_2}$ *se relacionan por*

$$\left(\vec{x}\right)_{\mathfrak{B}_2} = M_{\mathfrak{B}_2}^{-1} M_{\mathfrak{B}_1} \left(\vec{x}\right)_{\mathfrak{B}_1}. \tag{1.22}$$

Demostración.

Queda como ejercicio para el lector. ■

1.6.3 Espacios con producto interno

En esta sección se define una operación que no puede ser clasificada como externa o interna, pues aunque se opera con los elementos de un espacio vectorial, el resultado es un escalar el cual no pertenece al conjunto sobre el que se define la operación.

Definición 1.46 Espacio con producto interno

Un espacio vectorial real $\mathbb{V}$ *de dimensión finita se dice que es un espacio con producto interno si a cada par de vectores* $\vec{u}, \vec{v} \in \mathbb{V}$ *le asigna un número real denotado por* $\langle \vec{u}, \vec{v} \rangle$*, tal que*

$$\langle \vec{u}, \vec{v} \rangle = \vec{u} \cdot \vec{v} = \vec{u}\ {}^t\vec{v} = \sum_{i=1}^{n} u_i v_i. \tag{1.23}$$

Debido a la notación en (1.23)*, el producto interno se llama con frecuencia producto escalar o producto punto entre vectores.*

Teorema 1.32 Propiedades del producto interno

Sea $\mathbb{V}$ *un espacio vectorial real con un producto interno* $\langle\ ,\ \rangle$*. Entonces para todo* $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}$ *y todo* $\alpha \in \mathbb{R}$*, se tiene*

i) $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$.

ii) $\big\langle \big(\vec{u} + \vec{v}\big), \vec{w} \big\rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$.

iii) $\big\langle \vec{u}, \big(\vec{v} + \vec{w}\big) \big\rangle = \langle \vec{u}, \vec{v} \rangle + \langle \vec{u}, \vec{w} \rangle$.

iv) $\big\langle \big(\alpha\vec{u}\big), \vec{v} \big\rangle = \alpha \langle \vec{u}, \vec{v} \rangle = \big\langle \vec{u}, \big(\alpha\vec{v}\big) \big\rangle$.

v) $\langle \vec{u}, \vec{u} \rangle \geq 0$ *y* $\langle \vec{u}, \vec{u} \rangle = 0$ *si y solo si* $\vec{u} = \vec{0}$.

Demostración.

Queda como ejercicio para el lector. ■

Definición 1.47 Longitud o norma

Sea $\mathbb{V}$ un espacio vectorial real, con un producto interno $\langle\ ,\ \rangle$. Una norma en $\mathbb{V}$ es una función de $\mathbb{V}$ en $\mathbb{R}$, tal que a cada $\vec{v} \in \mathbb{V}$, le asigna un número real no negativo, denotado por $\|\vec{v}\|$ y definido como:

$$\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}. \tag{1.24}$$

Teorema 1.33 Propiedades de la norma

Para todo $\vec{u}, \vec{v} \in \mathbb{V}$ y todo $\alpha \in \mathbb{R}$,

(i) $\|\vec{u}\| \geq 0$.

(ii) $\|\vec{u}\| = 0$ *si y solo si* $\vec{u} = \vec{0}$.

(iii) $\|\alpha\vec{u}\| = |\alpha|\|\vec{u}\|$.

(iv) $\|\langle \vec{u}, \vec{v} \rangle\| \leq \|\vec{u}\|\|\vec{v}\|$.

(v) $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$.

Demostración.
Queda como ejercicio para el lector. ■

Teorema 1.34 *Sea $\langle\ ,\ \rangle$ un producto interno en un espacio vectorial $\mathbb{V}$ de dimensión finita. Sean $\vec{u}$ y $\vec{v}$ dos vectores diferentes de cero. Si θ es el ángulo entre ellos, entonces:*

$$\cos\theta = \cos \sphericalangle(\vec{u}, \vec{v}) = \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\|\|\vec{v}\|}. \tag{1.25}$$

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.48 Vectores ortogonales

1. *Sea $\mathbb{V}$ un espacio vectorial con un producto interno y sean dos vectores $\vec{u}, \vec{v} \in \mathbb{V}$. Se dice que $\vec{u}$ y $\vec{v}$ son ortogonales $\left(\vec{u} \perp \vec{v}\right)$ si y solo*

si:

$$\langle \vec{u}, \vec{v} \rangle = 0.$$

2. *Un conjunto de vectores* $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ *de un espacio vectorial* $\mathbb{V}$ *se dice que es ortogonal si y solo si los vectores son ortogonales dos a dos, es decir:*

$$\langle \vec{v}_i, \vec{v}_j \rangle = 0 \qquad \textit{siempre que} \qquad i \neq j.$$

3. *El conjunto de vectores* S *(en 2) se dice que es ortonormal si y solo si:*

$$a)\ S \textit{ es ortogonal.} \qquad b)\ \|\vec{v}_i\| = 1, \textit{ para todo } i.$$

Teorema 1.35 *Todo conjunto ortogonal de un espacio vectorial* $\mathbb{V}$ *es linealmente independiente.*

Demostración.
Queda como ejercicio para el lector. ■

Teorema 1.36 Proceso de ortonormalización de Gram-Schmidt

Todo subespacio $\mathbb{H}$ *de dimensión* k *de* $\mathbb{R}^n$ *tiene al menos una base ortogonal y una base ortonormal. Si* $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ *es cualquier base de* $\mathbb{H}$*, entonces:*

$$\begin{aligned}
\vec{w}_1 &= \vec{v}_1 \\
\vec{w}_2 &= \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 \\
\vec{w}_3 &= \vec{v}_3 - \frac{\langle \vec{v}_3, \vec{w}_1 \rangle}{\|\vec{w}_1\|^2}\vec{w}_1 - \frac{\langle \vec{v}_3, \vec{w}_2 \rangle}{\|\vec{w}_2\|^2}\vec{w}_2,
\end{aligned}$$

así sucesivamente:

$$\vec{w}_k \ = \vec{v}_k - \sum_{i=1}^{k-1} \frac{\langle \vec{v}_k, \vec{w}_i \rangle}{\|\vec{w}_i\|^2} \vec{w}_i$$

forman una base $\mathcal{B}' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ *ortogonal y*

$$\text{gen}\{\vec{v}_1, \ldots, \vec{v}_k\} \ = \ \text{gen}\{\vec{w}_1, \ldots, \vec{w}_k\}, \qquad i = 1, \ldots, k.$$

La base ortonormal $\mathcal{B}''$ *se obtiene normalizando* $\mathcal{B}'$*:*

$$\mathcal{B}'' = \left\{ \frac{\vec{w}_1}{\|\vec{w}_1\|}, \ldots, \frac{\vec{w}_k}{\|\vec{w}_k\|} \right\}.$$

Demostración.

Queda como ejercicio para el lector. ■

1.6.4 Complemento ortogonal

Consideremos un subespacio $\mathbb{V} \subseteq \mathbb{R}^n$. Para $\mathbb{V}$ puede haber muchos subespacios de $\mathbb{R}^n$ que son ortogonales a $\mathbb{V}$ (por ejemplo, en $\mathbb{R}^3$ si $\mathbb{V}$ es el eje Z, todas las rectas en el plano XY que pasan por el origen y el propio plano XY son ortogonales a $\mathbb{V}$). Entre todos los subespacios que son ortogonales a $\mathbb{V}$ hay uno de particular importancia: aquel subespacio $\mathbb{V}^*$ tal que $\mathbb{R}^n = \mathbb{V} \oplus \mathbb{V}^*$.

Definición 1.49 Complemento ortogonal

Sea $\mathbb{W}$ *un subespacio del espacio con producto interno* $\mathbb{V}$*. Entonces el complemento ortogonal de* $\mathbb{W}$*, denotado por* $\mathbb{W}^\perp$*, está dado por:*

$$\mathbb{W}^\perp = \left\{ \vec{u} \in \mathbb{V} : \langle \vec{u}, \vec{w} \rangle = 0, \ \textit{para todo } \vec{w} \in \mathbb{W} \right\}.$$

1.6.5 Subespacios asociados a una matriz

Hay cuatro subespacios asociados a una matriz, los cuales se consideran a continuación.

Definición 1.50 Espacio nulo y nulidad de una matriz

Sea A una matriz de tamaño $m \times n$. Entonces el conjunto

$$\ker(A) = \{\vec{u} \in \mathbb{R}^n | A\vec{u} = \vec{0}\} \tag{1.26}$$

se llama el espacio nulo de A y $\nu\big(A\big) = \dim\big[\ker(A)\big]$ se denomina nulidad de A.

El espacio nulo de una matriz también se conoce como *núcleo.*

Definición 1.51 Imagen de una matriz

Sea A una matriz de tamaño $m \times n$. Entonces la imagen de A, denotada por Im_A, está dada por:

$$Im_A = \{\vec{v} \in \mathbb{R}^m | A\vec{u} = \vec{v} \text{ para algún } \vec{u} \in \mathbb{R}^n\}. \tag{1.27}$$

Definición 1.52 Espacio de los renglones (o de las filas) de una matriz

Sea A una matriz de tamaño $m \times n$, sean $\{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m\}$ los renglones (o filas) de A. Entonces se define:

$$\mathcal{R}(A) = \text{espacio de los renglones de } A = \text{gen}\{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m\}. \tag{1.28}$$

Definición 1.53 Espacio de las columnas de una matriz

Sea A una matriz de tamaño $m \times n$, sean $\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ las columnas de A. Entonces se define:

$$Col(A) = \text{espacio de las columnas de } A = \text{gen}\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}. \tag{1.29}$$

Definición 1.54 Rango de una matriz

Sea $A = [a_{ij}]$ una matriz real de tamaño $m \times n$. Se llama rango de A, denotado por $\rho(A)$, al número máximo de vectores fila (o columna) linealmente independientes o también a la dimensión del subespacio generado por los vectores fila (o columna) de A.

Teorema 1.37 Propiedades del rango

Sea A una matriz real de tamaño $m \times n$, entonces se cumple que:

(i) $0 \leq \rho(A) \leq \text{mín}\{m, n\}$.

(ii) *Si $\rho(A) = m < n$, se dice que A tiene rango completo fila.*

(iii) *Si $\rho(A) = n < m$, se dice que A tiene rango completo columna.*

(iv) *Si $\rho(A) = r \leq \text{mín}\{m, n\}$, entonces existen matrices K y L de rango r y tamaños $m \times r$ y $r \times n$, respectivamente, tales que $A = KL$.*

(v) *$\rho(I_n) = n$, con I_n la matriz identidad de tamaño $n \times n$.*

(vi) *$\rho\big(\mathrm{O}\big) = 0$, con O la matriz nula de tamaño $n \times n$.*

(vii) $\rho(A) = \rho(A^t)$.

$(viii)$ $\rho(A.B) \leq \text{mín}\{\rho(A), \rho(B)\}$.

(ix) *Si A es diagonal, entonces $\rho(A)$ es el número de elementos no nulos en su diagonal.*

(x) *Si A es no singular, entonces $\rho(A.B) = \rho(B)$ y $\rho(B.A) = \rho(B)$.*

Demostración.

Queda como ejercicio para el lector. ■

1.7 Sistemas de ecuaciones lineales

En esta sección se mencionan las relaciones lineales, pues gran parte del álgebra lineal estudia y desarrolla relaciones lineales, las cuales son una generalización de la ecuación de una recta.

Definición 1.55 *Un sistema de ecuaciones se dice que es lineal si todas las ecuaciones que lo componen son lineales en los escalares desconocidos o incógnitas $x_1, x_2, \ldots, x_n$. Es decir, son de la forma:*

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_j x_j + \ldots + \alpha_n x_n = \beta,$$

donde α_i, $i = 1, 2, \ldots, n$ y β habitualmente son números reales, números complejos o funciones. Entonces, en una ecuación lineal no pueden aparecer productos o potencias de las incógnitas $x_1, x_2, \ldots, x_j, \ldots, x_n$.

Definición 1.56 *Se llama sistema lineal de m ecuaciones con n incógnitas al conjunto de m igualdades:*

$$b_i = \sum_{j=1}^{n} a_{ij} x_j \qquad \text{para} \qquad i = 1, 2, \ldots, m, \tag{1.30}$$

donde a_{ij}, $i = 1, 2, \ldots, m$ $j = 1, 2, \ldots, n$ son los coeficientes del sistema; b_i, $i = 1, 2, \ldots, m$ son los términos independientes, y $x_1, x_2, \ldots, x_j, \ldots, x_n$ son las incógnitas del sistema.

El sistema de ecuaciones (1.30) puede escribirse en forma matricial como sigue:

$$\underbrace{\begin{bmatrix} a_{11} & \ldots & a_{1j} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots & \ldots & \vdots \\ a_{i1} & \ldots & a_{ij} & \ldots & a_{in} \\ \vdots & \ddots & \vdots & \ldots & \vdots \\ a_{m1} & \ldots & a_{mj} & \ldots & a_{mn} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}}_{\vec{X}} = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_j \\ \vdots \\ b_n \end{bmatrix}}_{\vec{b}}, \tag{1.31}$$

donde A es la matriz de coeficientes de tamaño $m \times n$; $\vec{X} \in \mathbb{R}^n$ es el vector columna de las incógnitas y $\vec{b} \in \mathbb{R}^m$ es el vector columna de los términos independientes.

El sistema (1.31) se dice que es *homogéneo* cuando el vector $\vec{b}$ de términos independientes es nulo. Es decir,

$$A\vec{X} = \vec{0}.$$

Se conoce como *sistema lineal no homogéneo general* al sistema de la forma:

$$A\vec{X} = \vec{b} \qquad \text{con} \qquad \vec{b} \neq \vec{0}.$$

Definición 1.57 Sistemas consistentes e inconsistentes

Un sistema de ecuaciones lineales

$$A\vec{X} = \vec{b} \tag{1.32}$$

con A una matriz de tamaño $m \times n$, $\vec{X} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, se puede clasificar según el número de soluciones que tenga, como:

- *Inconsistente si no tiene solución.*
- *Consistente si admite al menos una solución. En este caso el sistema* (1.32) *puede clasificarse en:*
  - *Consistente determinado si la solución es única.*
  - *Consistente indeterminado cuando hay infinitas soluciones.*

1.7.1 Método de eliminación de Gauss

Para resolver un sistema de m ecuaciones con n incógnitas de la forma

$$A\vec{X} = \vec{b},$$

se procede de la siguiente manera:

1. Se forma la matriz aumentada $B = \left(A \mid \vec{b}\right)$.
2. Se aplican operaciones elementales entre filas, hasta llevar a B a una matriz escalonada reducida C.
3. Se halla el sistema de ecuaciones que representa la matriz C.
4. El último sistema determina la solución.

Por último, en esta sección se presenta la aplicación de determinantes, bien conocida para resolver sistemas de ecuaciones lineales $A\vec{X} = \vec{b}$, donde A es una matriz invertible de tamaño $n \times n$. Sean

$$A = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix} \qquad \text{y} \qquad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Para $j = 1, 2, \dots, n$, denotemos por B_j la matriz que resulta de sustituir la columna j-ésima de A por el vector $\vec{b}$:

$$B_j = \begin{bmatrix} a_{11} & \dots & a_{1,j-1} & b_1 & a_{1,j+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2,j-1} & b_2 & a_{2,j+1} & \dots & a_{2n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{n,j-1} & b_n & a_{n,j+1} & \dots & a_{nn} \end{bmatrix}. \tag{1.33}$$

Teorema 1.38 Regla de Cramer

Consideremos el sistema $A\vec{X} = \vec{b}$, donde A es invertible. Entonces, la solución del sistema viene dada por

$$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} \det B_1 \\ \det B_2 \\ \vdots \\ \det B_n \end{bmatrix}, \tag{1.34}$$

donde las matrices $B_1, B_2, \dots, B_n$, están definidas en (1.33).

Demostración.

Queda como ejercicio para el lector. ■

1.8 Transformaciones lineales

Definición 1.58 *Sean $\mathbb{V}$ y $\mathbb{W}$ espacios vectoriales reales. Sea $T : \mathbb{V} \to \mathbb{W}$ una función de $\mathbb{V}$ en $\mathbb{W}$. Se dice que T es una transformación lineal de $\mathbb{V}$ en $\mathbb{W}$ si y solo si para cualquier $\vec{u}$, $\vec{v}$ vectores de $\mathbb{V}$ y α escalar, se tiene que:*

$$T\big(\vec{u}+\vec{v}\big) = T\big(\vec{u}\big)+T\big(\vec{v}\big) \qquad y \qquad T\big(\alpha\vec{u}\big) = \alpha T\big(\vec{u}\big). \tag{1.35}$$

Teorema 1.39 *Sea T una transformación lineal de $\mathbb{V}$ en $\mathbb{W}$. Entonces:*

1. $T\big(\vec{0}\big) = \vec{0}$.
2. $T\big(-\vec{u}\big) = -T\big(\vec{u}\big)$, *para todo* $\vec{u} \in \mathbb{V}$.
3. *Si* $\vec{v} = \sum_{i=1}^{n} \alpha_i \vec{u}_i$, *con* $\vec{u}_i \in \mathbb{V}$ *y* α_i *escalares, entonces:*
$$T\left(\sum_{i=1}^{n} \alpha_i \vec{u}_i\right) = \sum_{i=1}^{n} \alpha_i T\big(\vec{u}_i\big).$$
4. *Si el conjunto* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ *es linealmente dependiente en* $\mathbb{V}$, *entonces el conjunto* $\big\{T\big(\vec{v}_1\big), T\big(\vec{v}_2\big), \ldots, T\big(\vec{v}_m\big)\big\}$ *es linealmente dependiente en* $\mathbb{W}$.

Demostración.
Queda como ejercicio para el lector. ■

1.8.1 Representación matricial de una transformación

En este apartado se verá que para toda transformación lineal $T : \mathbb{V} \to \mathbb{W}$, existe una matriz A de tamaño $m \times n$ con $m = \dim \mathbb{W}$ y $n = \dim \mathbb{V}$ tal que:

$$T\big(\vec{x}\big) = A\vec{x} \qquad \text{para todo} \qquad \vec{x} \in \mathbb{V}.$$

Este hecho es sumamente útil, ya que permite determinar de manera fácil el núcleo, la imagen, la nulidad y el rango de una transformación lineal.

Teorema 1.40 *Sean $\mathbb{V}$ y $\mathbb{W}$ espacios vectoriales reales de dimensiones n y m, respectivamente. Sea $T : \mathbb{V} \to \mathbb{W}$ una transformación lineal. Entonces existe una matriz única de tamaño $m \times n$, A_T tal que:*

$$T(\vec{x}) = A_T\vec{x} \qquad \text{para todo} \qquad \vec{x} \in \mathbb{V}.$$

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.59 Matriz de transformación

La matriz A_T en el Teorema 1.40 se llama matriz de transformación correspondiente a T o representación matricial de T.

Teorema 1.41 *Sea A_T la matriz de transformación correspondiente a la transformación lineal T. Entonces:*

i) $Im_T = Im_{A_T} = Col(A_T)$. *ii)* $\rho(T) = \rho(A_T)$.

iii) $\ker(T) = \ker(A_T)$. *iv)* $\nu(T) = \nu(A_T)$.

Demostración.
Queda como ejercicio para el lector. ■

Definición 1.60 Transformación uno a uno

Dada una transformación lineal $T : \mathbb{V} \to \mathbb{W}$, se dice que es uno a uno si satisface que:

$$T(\vec{v}_1) = T(\vec{v}_2) \qquad \text{implica que} \qquad \vec{v}_1 = \vec{v}_2. \tag{1.36}$$

Es decir, T es uno a uno (escrito $1-1$) si y solo si todo vector $\vec{v}$ en la imagen de T es la imagen de exactamente un vector en $\mathbb{V}$.

Definición 1.61 Transformación sobre

Sea $T : \mathbb{V} \to \mathbb{W}$ una transformación lineal. Entonces T es sobre, si para todo $\vec{w} \in \mathbb{W}$ existe al menos una $\vec{v} \in \mathbb{V}$ tal que $T(\vec{v}) = \vec{w}$. Es decir, T es sobre si y solo si $Im_T = \mathbb{W}$.

Definición 1.62 Isomorfismo

Sea $T : \mathbb{V} \to \mathbb{W}$ una transformación lineal. Entonces T es un isomorfismo si T es uno a uno y sobre.

1.9 Matrices con entradas complejas

En esta sección se desarrollarán algunas de las propiedades de las matrices cuyos elementos son números complejos. Toda la aritmética y los teoremas que se han expuesto se aplican a matrices complejas. Estas matrices tienen importantes aplicaciones, por ejemplo, en la mecánica cuántica.

1.9.1 Definición y propiedades básicas

Definición 1.63 Matriz compleja

Una matriz A de tamaño $m \times n$ se dice que es una matriz compleja si sus elementos son números complejos.

Definición 1.64 Matriz conjugada

Sea $A = [a_{ij}]$ una matriz compleja, se llama matriz conjugada de A a la matriz $\overline{A} = [\overline{a}_{ij}]$, donde $\overline{a}_{ij}$ es el conjugado complejo de a_{ij}.

Ejemplo 1.9 *Determine la matriz conjugada de la matriz compleja*

$$A = \begin{bmatrix} 2-3i & 2+i \\ 2-i & 2i \end{bmatrix}.$$

Solución.

$$\overline{A} = \begin{bmatrix} \overline{2-3i} & \overline{2+i} \\ \overline{2-i} & \overline{2i} \end{bmatrix} = \begin{bmatrix} 2+3i & 2-i \\ 2+i & -2i \end{bmatrix}.$$

Teorema 1.42 Propiedades de la conjugada compleja

Sean A y B matrices de componentes complejas de tamaño $m \times n$ y sea $\alpha \in \mathbb{C}$. Entonces:

1. $\overline{A+B} = \overline{A} + \overline{B}$.
2. $\overline{\overline{A}} = A$.
3. $\overline{A^t} = \overline{A}^t$.
4. $\overline{\alpha A} = \overline{\alpha}\overline{A}$.
5. $\overline{AB} = \overline{A}\ \overline{B}, \quad m = n$.

Demostración.

Sean $A = [a_{ij}]$ y $B = [b_{ij}]$, entonces

1. $\overline{A+B} = \left[\overline{a_{ij}+b_{ij}}\right] = \left[\overline{a_{ij}} + \overline{b_{ij}}\right] = \overline{A} + \overline{B}$.
2. $\overline{\overline{A}} = [\overline{\overline{a_{ij}}}] = [a_{ij}] = A$.
3. Queda como ejercicio para el lector.
4. $\overline{\alpha A} = [\overline{\alpha a_{ij}}] = [\overline{\alpha}\,\overline{a_{ij}}] = \overline{\alpha}\overline{A}$.
5. Definamos $C = AB$, luego el conjugado del elemento c_{ik} es

$$\begin{aligned} \overline{c_{ik}} &= \overline{a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk}} \\ &= \overline{\sum_{j=1}^{n} a_{ij}b_{jk}} = \sum_{j=1}^{n} \overline{a_{ij}b_{jk}} = \sum_{j=1}^{n} \overline{a_{ij}}\ \overline{b_{jk}} \\ &= \overline{a_{i1}}\,\overline{b_{1k}} + \overline{a_{i2}}\,\overline{b_{2k}} + \ldots + \overline{a_{in}}\,\overline{b_{nk}}. \end{aligned}$$

■

Definición 1.65 Transpuesta conjugada *La transpuesta conjugada de una matriz compleja A, denotada por A^H, se define como:*

$$A^H = \overline{A^t}, \tag{1.37}$$

donde los elementos de $\overline{A}$ son los conjugados complejos de los elementos correspondientes de A.

Ejemplo 1.10 *Determine A^H para la matriz*

$$A = \begin{bmatrix} 4+3i & 2+i \\ 2-i & 6i \\ -1 & 1+3i \end{bmatrix}.$$

Solución.

$$\overline{A} = \begin{bmatrix} \overline{4+3i} & \overline{2+i} \\ \overline{2-i} & \overline{6i} \\ \overline{-1} & \overline{1+3i} \end{bmatrix} = \begin{bmatrix} 4-3i & 2-i \\ 2+i & -6i \\ -1 & 1-3i \end{bmatrix}$$

$$A^H = \overline{A}^{\,t} = \begin{bmatrix} 4-3i & 2+i & -1 \\ 2-i & -6i & 1-3i \end{bmatrix}.$$

Teorema 1.43 Propiedades de la transpuesta conjugada

Si A y B son matrices complejas de tamaño $m \times n$ y $\alpha \in \mathbb{C}$, entonces se cumplen las siguientes propiedades:

1. $(A^H)^H = A$.
2. $(A+B)^H = A^H + B^H$.
3. $(\alpha\ A)^H = \overline{\alpha}\ A^H$.
4. $(A\ B)^H = B^H\ A^H, \quad m = n$.

Demostración.

Queda como ejercicio para el lector. ■

Teorema 1.44 *Sea A una matriz compleja de tamaño $n \times n$, entonces:*

$$\det(\overline{A}) = \overline{\det(A)}.$$

Demostración.

La prueba se hará por inducción sobre n. Sea $A = [a]$ una matriz de tamaño 1×1, entonces es claro que

$$\overline{\det(A)} = \overline{a} = \det(\overline{A}).$$

Ahora, supongamos que el teorema es cierto para matrices de tamaño $(n-1) \times (n-1)$.

Sea $A = [a_{ij}]$ una matriz de tamaño $n \times n$. Si se calcula el $\det(A)$ por la k-ésima fila, se tiene que

$$\det(A) = \sum_{j=1}^{n} a_{kj} C_{kj}(A) = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} M_{kj}(A),$$

donde $M_{kj}(A)$ es el menor complementario $\big(k, j\big)$. Por la hipótesis de inducción, se verifica que

$$\overline{M_{kj}(A)} = M_{kj}(\overline{A}).$$

Por lo tanto,

$$\begin{aligned}
\overline{\det(A)} &= \overline{\sum_{j=1}^{n} (-1)^{k+j} a_{kj} M_{kj}(A)} = \sum_{j=1}^{n} (-1)^{k+j} \overline{a_{kj} M_{kj}(A)} \\
&= \sum_{j=1}^{n} (-1)^{k+j} \overline{a}_{kj} \overline{M_{kj}(A)} = \sum_{j=1}^{n} (-1)^{k+j} \overline{a}_{kj} M_{kj}(\overline{A}) \\
&= \det(\overline{A}). \quad \blacksquare
\end{aligned}$$

1.9.2 Espacios vectoriales complejos

Las propiedades algebraicas de un *espacio vectorial complejo* arbitrario son muy semejantes a las estudiadas para los *espacios vectoriales reales*.

Definición 1.66 Espacios vectoriales complejos

Un espacio vectorial complejo se define exactamente como un espacio vectorial real (definición 1.36), excepto que los escalares en los axiomas (vi) a (ix) pueden ser números complejos. Los términos espacio vectorial complejo y espacio vectorial real destacan el conjunto del cual se eligen los escalares.

Los conceptos de combinaciones lineales, conjuntos generadores, dependencia lineal, independencia lineal y base no cambian para los espacios vectoriales complejos, excepto que utilizamos escalares complejos.

Definición 1.67 Producto interno en $\mathbb{C}^n$

Sean $\vec{u}, \vec{v} \in \mathbb{C}^n$, se define el producto punto $\vec{u} \cdot \vec{v}$ como

$$\vec{u} \cdot \vec{v} = \vec{u}^H \vec{v} = \sum_{i=1}^{n} \overline{u_i} v_i,$$

donde $\overline{u_i}$ es el i-ésimo elemento conjugado de $\vec{u}$ y v_i es el i-ésimo elemento de $\vec{v}$.

Teorema 1.45 Propiedades del producto punto en $\mathbb{C}^n$

Para todo $\vec{u}, \vec{v}, \vec{w} \in \mathbb{C}^n$ y todo $\alpha \in \mathbb{C}$:

i) $\vec{u} \cdot \vec{u} \geq 0$.

ii) $\vec{u} \cdot \vec{u} = 0$ *si y solo si* $\vec{u} = \vec{0}$.

iii) $\vec{u} \cdot \vec{v} = \overline{\vec{v} \cdot \vec{u}}$.

iv) $\big(\vec{u} + \vec{v}\big) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$.

v) $\big(\alpha \vec{u}\big) \cdot \vec{v} = \overline{\alpha}\big(\vec{u} \cdot \vec{v}\big)$.

Demostración.

Queda como ejercicio para el lector. ■

Definición 1.68 Partes real e imaginaria de un vector complejo

El complejo conjugado de un vector complejo $\vec{u} \in \mathbb{C}^n$ *es el vector* $\overline{\vec{u}} \in \mathbb{C}^n$*, cuyas componentes son los complejos conjugados de las componentes de* $\vec{u}$*. Las partes real e imaginaria de un vector complejo* $\vec{u}$ *son los vectores* $Re(\vec{u}) \in \mathbb{R}^n$ *y* $Im(\vec{u}) \in \mathbb{R}^n$*, respectivamente, formados a partir de las partes reales e imaginarias de cada una de las componentes de* $\vec{u}$*.*

Ejemplo 1.11 *Determine las partes real e imaginaria y el correspondiente vector conjugado del vector* $\vec{u}^{\,t} = (-i, 1+i, 1)$*.*

Solución.

Como

$$\vec{u} = \begin{bmatrix} -i \\ 1+i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + i \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

entonces

$$Re(\vec{u}) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \qquad \text{y} \qquad Im(\vec{u}) = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

Luego, el vector conjugado es

$$\overline{\vec{u}} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - i \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} i \\ 1-i \\ 1 \end{bmatrix}.$$

1.9.3 Solución de sistemas lineales con entradas complejas

Los resultados y las técnicas para resolver sistemas lineales, presentados en la sección 1.7, se pueden aplicar de manera directa a los sistemas lineales con coeficientes complejos. En este apartado, se muestra cómo transformar un sistema lineal de $n \times n$ con coeficientes complejos en un sistema lineal $2n \times 2n$ con coeficientes reales.

Consideremos el sistema

$$A\vec{X} = \vec{b}, \tag{1.38}$$

donde A es una matriz compleja de tamaño $n \times n$, $\vec{X}, \vec{b} \in \mathbb{C}^n$. Entonces, el sistema dado en (1.38) se puede escribir como

$$\begin{bmatrix} Re(A) & -Im(A) \\ Im(A) & Re(A) \end{bmatrix} \begin{bmatrix} Re(\vec{x}) \\ Im(\vec{x}) \end{bmatrix} = \begin{bmatrix} Re(\vec{b}) \\ Im(\vec{b}) \end{bmatrix},$$

donde $Re(\cdot)$ y $Im(\cdot)$ denotan las partes real e imaginaria, respectivamente. Este nuevo sistema lineal con coeficientes reales es de $2n$ ecuaciones con $2n$ incógnitas. Si se emplean los resultados de la sección 1.5, se tiene que el sistema dado en (1.38) tiene una única solución si y solo si

$$\det\{Re(A)\} \neq 0 \quad \text{y} \quad \det\Big\{Re(A) + Im(A)\big[Re(A)\big]^{-1} Im(A)\Big\} \neq 0.$$

En cuyo caso la solución está dada por

$$\begin{bmatrix} Re(\vec{x}) \\ Im(\vec{x}) \end{bmatrix} = \begin{bmatrix} I_n & \big[Re(A)\big]^{-1} Im(A) \\ -\big[Re(A)\big]^{-1} Im(A) & I_n \end{bmatrix} \begin{bmatrix} C^{-1} Re(\vec{b}) \\ C^{-1} Im(\vec{b}) \end{bmatrix},$$

donde I_n es la matriz identidad de tamaño $n \times n$ y la matriz

$$C = Re(A) + Im(A)\big[Re(A)\big]^{-1} Im(A).$$

Ejemplo 1.12 *Determine una solución del sistema de ecuaciones:*

$$\begin{aligned} (2-i)x_1 + (1-i)x_2 &= 3 - 6i \\ (3+i)x_1 + (2+2i)x_2 &= 7 + i. \end{aligned} \tag{1.39}$$

Solución.

Al expresar matricialmente (1.39), se llega a

$$\begin{bmatrix} 2-i & 1-i \\ 3+i & 2+2i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3-6i \\ 7+i \end{bmatrix}. \tag{1.40}$$

Como

$$A = \begin{bmatrix} 2-i & 1-i \\ 3+i & 2+2i \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} + i \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix} = Re(A) + i\, Im(A).$$

De manera análoga, el término independiente se puede expresar

$$\begin{bmatrix} 3-6i \\ 7+i \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix} + i \begin{bmatrix} -6 \\ 1 \end{bmatrix}.$$

Si $x_1 = a_1 + ib_1$ y $x_2 = a_2 + ib_2$, entonces (1.39) se puede reescribir como

$$\begin{bmatrix} Re(A) & \vdots & -Im(A) \\ \cdots & . & \cdots \\ Im(A) & \vdots & Re(A) \end{bmatrix} \begin{bmatrix} Re(\vec{x}) \\ \cdots \\ Im(\vec{x}) \end{bmatrix} = \begin{bmatrix} Re(\vec{b}) \\ \cdots \\ Im(\vec{b}) \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & \vdots & 1 & 1 \\ 3 & 2 & \vdots & -1 & -2 \\ \cdots & \cdots & . & \cdots & \cdots \\ -1 & -1 & \vdots & 2 & 1 \\ 1 & 2 & \vdots & 3 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ \cdots \\ -6 \\ 1 \end{bmatrix}.$$

Para hallar C, se realizan los respectivos productos y se llega a

$$C = \begin{bmatrix} 0 & -2 \\ 10 & 12 \end{bmatrix} \qquad \text{y} \qquad \det C = 20 \neq 0.$$

Por otra parte, el determinante de $Re(A) = 1 \neq 0$. Luego, el sistema tiene única solución y está dada por

$$\begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \vdots & -3 & -4 \\ 0 & 1 & \vdots & 5 & 7 \\ \cdots & \cdots & . & \cdots & \cdots \\ 3 & 4 & \vdots & 1 & 0 \\ -5 & -7 & \vdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{5}{2} \\ -\frac{3}{2} \\ \cdots \\ -\frac{7}{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ \cdots \\ -2 \\ 1 \end{bmatrix}.$$

Por tanto, la solución del sistema lineal dado es

$$x_1 = 1 - 2i \qquad \text{y} \qquad x_2 = 2 + i.$$

Capítulo 2

Vectores característicos y valores característicos

En una gran variedad de aplicaciones, dada una transformación lineal $T : \mathbb{V} \to \mathbb{V}$, resulta útil encontrar un vector $\vec{v}$ en $\mathbb{V}$ tal que $T\vec{v}$ y $\vec{v}$ sean paralelos. Esto es, se busca un vector $\vec{v}$ y un escalar λ tal que

$$T\vec{v} = \lambda\vec{v} \tag{2.1}$$

tenga una solución $\vec{v} \neq \vec{0}$. En este caso, λ se denomina *valor característico* de T y $\vec{v}$ se llama *vector característico* de T correspondiente al valor característico λ. Si $\dim(\mathbb{V}) = n$, el problema de determinar los respectivos valores característicos de T puede resolverse con la ayuda de los determinantes. Nótese que la ecuación (2.1) puede escribirse en la forma

$$\big(T - \lambda I_n\big)\vec{v} = \vec{0},$$

donde I_n es la transformación identidad. Si denotamos $T_\lambda = T - \lambda I_n$, entonces λ es un valor característico si y solo si la ecuación

$$T_\lambda(\vec{v}) = \vec{0} \tag{2.2}$$

tiene una solución $\vec{v}$ no nula, en cuyo caso T_λ no es invertible, pues una solución no nula de (2.2) existe si y solo si la matriz de T_λ es singular. Si A_T es una representación matricial de T, entonces $A_T - \lambda I_n$ es una representación matricial para T_λ. Por esta razón, en este capítulo se estudiarán algunas de las propiedades de los valores y vectores característicos de las matrices de tamaño $n \times n$.

2.1 Valores propios y vectores propios

Definición 2.1 Valor característico y vector característico

Un vector característico de una matriz A de tamaño $n \times n$ es un vector $\vec{v}$ diferente de cero, que cumple:

$$A\vec{v} = \lambda\vec{v} \tag{2.3}$$

para algún escalar λ. El escalar λ es llamado valor característico de A si existe una solución no trivial $\vec{v}$ del sistema $A\vec{v} = \lambda\vec{v}$. También, $\vec{v}$ se denomina vector característico correspondiente a λ.

Nota 2.1 Los valores característicos se llaman también *autovalores, valores propios o eigenvalores*, y los vectores característicos, *autovectores, vectores propios o eigenvectores.*

Teorema 2.1 *Sea $\vec{v}$ un vector propio de una matriz A asociado al valor propio λ. Sea $\alpha \neq 0$, entonces $\alpha\vec{v}$ también es un vector propio de A correspondiente al valor propio λ.*

Demostración.

Se debe probar que $\alpha\vec{v}$ satisface (2.3). Utilizando el hecho de que $A\vec{v} = \lambda\vec{v}$, se tiene que

$$A(\alpha\vec{v}) = \alpha(A\vec{v}) = \alpha(\lambda\vec{v}) = \lambda(\alpha\vec{v}),$$

lo cual completa la prueba. ■

Ejemplo 2.1 *Verifique si los vectores $\vec{u}^{\,t} = (-1, 1)$ y $\vec{v}^{\,t} = (2, 1)$ son vectores propios de la siguiente matriz:*

$$A = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix}.$$

Solución.

Se tiene que

$$A\vec{u} = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1\vec{u}.$$
$$A\vec{v} = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 14 \end{bmatrix} \neq \lambda \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Así, $\vec{u}$ es un vector propio correspondiente al valor propio $\lambda = -1$, pero $\vec{v}$ no es un vector propio de A porque $A\vec{v}$ no es un múltiplo de $\vec{v}$, es decir no existe un escalar λ tal que $2\lambda = 4$ y $\lambda = 14$ se verifiquen simultáneamente.

Ejemplo 2.2 *Considere la matriz A dada en el Ejemplo 2.1, muestre que 6 es un valor propio de A y encuentre el vector propio correspondiente.*

Solución.

El escalar 6 es un valor propio de A si y solo si la ecuación

$$A\vec{v} = 6\vec{v}, \tag{2.4}$$

tiene una solución no trivial. Pero (2.4) es equivalente a $A\vec{v} - 6\vec{v} = \vec{0}$, o

$$\big(A - 6I\big)\vec{v} = \vec{0} \tag{2.5}$$

Para resolver esta ecuación homogénea, se forma la matriz

$$A - 6I = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix} - \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} -5 & 2 \\ 5 & -2 \end{bmatrix}.$$

Las columnas de $A - 6I$ son linealmente dependientes, es decir (2.4) tiene solución no trivial; luego, 6 es un valor propio de A. Para encontrar el vector propio correspondiente, se realizan operaciones por fila

$$\left[\begin{array}{cc|c} -5 & 2 & 0 \\ 5 & -2 & 0 \end{array}\right] \underset{F_2 + F_1}{\sim} \left[\begin{array}{cc|c} -5 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

La solución general tiene la forma $\begin{bmatrix} 1 \\ \frac{5}{2} \end{bmatrix} x$. Según el Teorema 2.1, cada vector de esta forma con $x \neq 0$ es un vector propio correspondiente a $\lambda = 6$.

Teorema 2.2 *Sea $A = [a_{ij}]$ una matriz real de tamaño $n \times n$ e I_n la matriz identidad de tamaño $n \times n$, entonces la función $p_A(\lambda)$ es definida por la ecuación*

$$p_A(\lambda) = \det(A - \lambda I_n) \; = \; \sum_{k=0}^{n} \mathrm{tr}_k(A) \left(-\lambda\right)^{n-k}, \tag{2.6}$$

donde $\mathrm{tr}_k(A)$ denota la suma de los determinantes de las submatrices principales de orden k, $\left[\mathrm{tr}_0(A) = 1 \text{ y } \mathrm{tr}_n(A) = \det(A)\right]$, es un polinomio en λ de grado n y el término independiente es $p_A(0) = \det(A)$.

Demostración.

Vamos a demostrar que $p_A(\lambda)$ es un polinomio de grado n únicamente para el caso $n \leq 3$. La demostración para el caso general puede hacerse por inducción.

Para $n = 1$, el determinante es el polinomio lineal $p_A(\lambda) = a_{11} - \lambda$. Para $n = 2$, se tiene que

$$\begin{aligned} p_A(\lambda) \; &= \; \det(A - \lambda I_2) = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ &= \; \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = \lambda^2 - \mathrm{tr}_1(A)\lambda + \det(A). \end{aligned}$$

Obsérvese que el polinomio obtenido es de segundo grado en λ. Para $n = 3$ tenemos

$$\begin{aligned} p_A(\lambda) \; &= \; \det(A - \lambda I_3) = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} \\ &= (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - [a_{32}a_{23}(a_{11} - \lambda) + a_{13}a_{31}(a_{22} - \lambda) + a_{21}a_{12}(a_{33} - \lambda)] \\ &= \; -\lambda^3 + \mathrm{tr}_1(A)\lambda^2 - \big(\sum_{i=1}^{3} M_{ii}(A)\big)\lambda + \det(A), \end{aligned}$$

donde $M_{ii}(A)$ denota el menor complementario (i, i) de A. Nótese que en este caso se obtiene un polinomio de tercer grado, siendo el término de mayor grado $-\lambda^3$.

La afirmación $p_A(0) = \det(A)$ resulta inmediata de la definición de p_A.

■

Teorema 2.3 *Sea A una matriz real de tamaño $n \times n$. Entonces λ es un valor propio de A si y solo si*

$$p_A(\lambda) = \det\big(A - \lambda I_n\big) = 0. \tag{2.7}$$

Demostración.

Supóngase que λ es un valor propio de A. Entonces existe un elemento $\vec{v} \neq \vec{0}$ tal que $A\vec{v} = \lambda\vec{v}$, de donde $A\vec{v} - \lambda\vec{v} = \vec{0}$ o $\big(A - \lambda I_n\big)\vec{v} = \vec{0}$. Por lo tanto, $A - \lambda I_n$ tiene un núcleo no nulo y $A - \lambda I_n$ no es invertible, es decir $\det\big(A - \lambda I_n\big) = 0$.

Recíprocamente, supóngase que $\det\big(A - \lambda I_n\big) = 0$, es decir $A - \lambda I_n$ no es invertible. Entonces $A - \lambda I_n$ debe tener un núcleo no nulo, lo que significa que existe un elemento $\vec{v} \neq \vec{0}$ tal que $\big(A - \lambda I_n\big)\vec{v} = \vec{0}$. Por lo tanto, $A\vec{v} - \lambda\vec{v} = \vec{0}$ o $A\vec{v} = \lambda\vec{v}$. Así, λ es un valor propio de A. ■

Definición 2.2 Ecuación y polinomio característico

La ecuación (2.6) *se llama polinomio característico de A. La ecuación* (2.7) *se llama ecuación característica de A.*

Definición 2.3 Multiplicidad algebraica

Sea λ_k un valor propio de una matriz A de tamaño $n \times n$. Entonces, la multiplicidad algebraica de λ_k es el número de veces que λ_k aparece como raíz del polinomio característico de A; es decir, es igual a su multiplicidad como raíz de la ecuación característica.

Ejemplo 2.3 *Encuentre el polinomio y la ecuación característica de*

$$A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}.$$

Solución.

Fórmese $A - \lambda I$ y calcúlese su determinante

$$\begin{aligned}\det(A-\lambda I) &= \begin{vmatrix} 4-\lambda & -1 & 6 \\ 2 & 1-\lambda & 6 \\ 2 & -1 & 8-\lambda \end{vmatrix} \\ &= (4-\lambda)\begin{vmatrix} 1-\lambda & 6 \\ -1 & 8-\lambda \end{vmatrix} - (-1)\begin{vmatrix} 2 & 6 \\ 2 & 8-\lambda \end{vmatrix} + 6\begin{vmatrix} 2 & 1-\lambda \\ 2 & -1 \end{vmatrix} \\ &= (4-\lambda)[(1-\lambda)(8-\lambda)+6] + [2(8-\lambda)-12] + 6[-2-2(1-\lambda)].\end{aligned}$$

Simplificando el producto, se obtiene el polinomio característico

$$p_A(\lambda) = -\lambda^3 + 13\lambda^2 - 40\lambda + 36.$$

Los resultados del Teorema 2.2 se cumplen, ya que la $\mathrm{tr}_1(A) = 13$, $\sum_{i=1}^{3} M_{ii}(A) = 40$ y el $\det(A) = 36$, y factorizando p_A, se llega a

$$p_A(\lambda) = -(\lambda-2)^2(\lambda-9).$$

En este caso, la ecuación característica es

$$(\lambda-2)^2(\lambda-9) = 0.$$

Nótese que el valor propio 2 tiene *multiplicidad algebraica* 2 pues $(\lambda - 2)$ aparece dos veces como factor del polinomio característico.

Teorema 2.4 *Sea A una matriz real de tamaño $n \times n$ con valores propios distintos $\lambda_1, \lambda_2, \ldots, \lambda_m$. Sea*

$$\mathcal{B}_k = \{\vec{v} \in \mathbb{C}^n : A\vec{v} = \lambda_k \vec{v}\} \qquad \textit{para} \qquad k = 1, 2, \ldots, m, \tag{2.8}$$

entonces para cada k, $\mathcal{B}_k = \ker\{A - \lambda_k I\}$ es un subespacio de $\mathbb{C}^n$.

Demostración.

Como $A\vec{0} = \lambda_k \vec{0}$ para todo k, $\vec{0} \in \mathcal{B}_k$. Si λ_k no es un valor propio, no existen vectores $\vec{v} \neq \vec{0}$ excepto $\vec{0}$ que satisface $A\vec{v} = \lambda_k \vec{v}$. En este caso, $\mathcal{B}_k$ es el *subespacio trivial*.

Ahora supongamos que λ_k es un valor propio. Entonces existe un $\vec{v} \neq \vec{0}$ tal que $A\vec{v} = \lambda_k \vec{v}$, en otras palabras $(A - \lambda_k I)\vec{v} = \vec{0}$. De esta manera, $\mathcal{B}_k = \{\vec{v} \in \mathbb{C}^n : (A - \lambda_k I)\vec{v} = \vec{0}\}$ es el *espacio solución* del sistema homogéneo $(A - \lambda_k I)\vec{v} = \vec{0}$, el cual es un subespacio de $\mathbb{C}^n$. ■

Definición 2.4 Espacio propio

Sea λ_k un valor propio de A. El subespacio $\mathcal{B}_k$ se denomina espacio propio de A correspondiente al valor propio λ_k.

Teorema 2.5 *Sea A una matriz real de tamaño $n \times n$. Entonces los espacios propios correspondientes a valores propios distintos de A tienen al vector nulo en común. Es decir, $\mathcal{B}_k \cap \mathcal{B}_r = \left\{\vec{0}\right\}$ si $\lambda_k \neq \lambda_r$.*

Demostración.

Supongamos que existe un vector propio de A tal que $\vec{v} \in \mathcal{B}_k$ y $\vec{v} \in \mathcal{B}_r$, entonces por (2.8) se tiene que

$$A\vec{v} = \lambda_k \vec{v} \qquad \text{y} \qquad A\vec{v} = \lambda_r \vec{v}.$$

Luego,

$$(\lambda_k - \lambda_r)\vec{v} = \vec{0},$$

pero como $\lambda_k \neq \lambda_r$, se concluye que $\vec{v} = \vec{0}$; por lo tanto, $\mathcal{B}_k \cap \mathcal{B}_r = \left\{\vec{0}\right\}$ y la prueba queda completa. ■

Teorema 2.6 *Si $\vec{v}_1, \ldots, \vec{v}_m$ son vectores propios correspondientes a valores propios distintos $\lambda_1, \ldots, \lambda_m$ de una matriz A de tamaño $n \times n$, entonces el conjunto $\{\vec{v}_1, \ldots, \vec{v}_m\}$ es linealmente independiente.*

Demostración.

La demostración es por inducción sobre m. El resultado es trivial cuando $m = 1$. Entonces supongamos que se ha demostrado para cualquier conjunto $m = k$ de vectores propios. Sean $\vec{v}_1, \ldots, \vec{v}_{k+1}$, $k+1$ vectores propios pertenecientes a valores propios distintos y supongamos que existen escalares c_i tales que

$$\sum_{i=1}^{k+1} c_i \vec{v}_i = \vec{0}. \tag{2.9}$$

Si se multiplica por A a ambos lados de (2.9) y se utiliza el hecho de que $A\vec{v}_i = \lambda_i \vec{v}_i$, se obtiene

$$\sum_{i=1}^{k+1} c_i \lambda_i \vec{v}_i = \vec{0}. \tag{2.10}$$

Al restar de (2.10) el producto de (2.9) por λ_{k+1}, se obtiene la ecuación

$$\sum_{i=1}^{k} c_i (\lambda_i - \lambda_{k+1}) \vec{v}_i = \vec{0}.$$

Como los vectores $\vec{v}_1, \ldots, \vec{v}_k$ son linealmente independientes por hipótesis de inducción, se debe tener que $c_i(\lambda_i - \lambda_{k+1}) = 0$ para cada $i = 1, 2, \ldots, k$. Además, como los valores propios son distintos, se tiene que $\lambda_i \neq \lambda_{k+1}$ para $i \neq k+1$, así que $c_i = 0$ para $i = 1, 2, \ldots, k$, y de (2.9) se tiene que c_{k+1} es también 0. Por lo tanto, el conjunto de vectores propios $\vec{v}_1, \ldots, \vec{v}_{k+1}$ es también linealmente independiente. ■

Teorema 2.7 *Los vectores no nulos tomados de espacios propios distintos son linealmente independientes. En otras palabras, los espacios propios $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_m$ (correspondientes a los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_m$ distintos) cumplen que*

"Si $\vec{u}_1 + \ldots + \vec{u}_m = \vec{0}$ con $\vec{u}_k \in \mathcal{B}_k$ entonces $\vec{u}_1 = \ldots = \vec{u}_m = \vec{0}$."

Demostración.

Supongamos que $\vec{u}_1 + \ldots + \vec{u}_m = \vec{0}$ con $\vec{u}_k \in \mathcal{B}_k$, esto es $A\vec{u}_k = \lambda_k \vec{u}_k$. Si algunos $\vec{u}_k$ no fueran $\vec{0}$, ellos serían vectores propios de A, correspondientes a valores propios distintos, entonces el que la suma de ellos sea $\vec{0}$ contradice el Teorema 2.6. ■

Definición 2.5 Multiplicidad geométrica

Sea λ_k un valor propio de una matriz A de tamaño $n \times n$. Entonces la multiplicidad geométrica de λ_k es el número máximo de vectores propios de A linealmente independientes que tienen un valor propio igual a λ_k. Es decir, es igual a la dimensión del espacio propio correspondiente a λ_k (lo cual es la nulidad de la matriz $A - \lambda_k I$). En consecuencia,

$$\textit{multiplicidad geométrica de } \lambda_k = \dim(\mathcal{B}_k) = \nu\big(A - \lambda_k I_n\big).$$

Nota 2.2 La multiplicidad geométrica de un valor propio nunca es cero. Esto se establece de la Definición 2.1, la cual expresa que si λ es un valor propio, entonces existe un vector propio *no nulo* correspondiente a λ.

A continuación, se presenta un procedimiento para calcular los valores propios y vectores propios de una matriz A de tamaño $n \times n$.

Determinación de valores propios y vectores propios

i) Encuentre $p_A(\lambda) = \det(A - \lambda I)$.

ii) Halle las raíces $\lambda_1, \lambda_2, \ldots, \lambda_m$ de $p_A(\lambda) = 0$.

iii) Resuelva el sistema homogéneo $(A - \lambda_k I)\vec{v} = \vec{0}$, correspondiente a cada valor propio λ_k.

Ejemplo 2.4 *Encuentre los vectores propios y espacios propios asociados a la matriz dada en el Ejemplo 2.3.*

Solución.

En el Ejemplo 2.3, se obtuvo que la ecuación característica era

$$(\lambda - 2)^2(\lambda - 9) = 0.$$

De esta manera, los valores propios de A son $\lambda_1 = 9$ y $\lambda_2 = 2$ (con multiplicidad algebraica 2).

Para $\lambda_1 = 9$, se tiene

$$(A - 9I)\vec{v} = \begin{bmatrix} -5 & -1 & 6 \\ 2 & -8 & 6 \\ 2 & -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Para determinar el vector propio correspondiente, se realizan operaciones por filas

$$\left[\begin{array}{ccc|c} -5 & -1 & 6 & 0 \\ 2 & -8 & 6 & 0 \\ 2 & -1 & -1 & 0 \end{array}\right] \underset{F_1 \leftrightarrow \frac{1}{2}F_2}{\sim} \left[\begin{array}{ccc|c} 1 & -4 & 3 & 0 \\ -5 & -1 & 6 & 0 \\ 2 & -1 & -1 & 0 \end{array}\right] \underset{\substack{F_2 + 5F_1 \\ F_3 - 2F_1}}{\sim} \left[\begin{array}{ccc|c} 1 & -4 & 3 & 0 \\ 0 & -21 & 21 & 0 \\ 0 & 7 & -7 & 0 \end{array}\right]$$

$$\underset{\substack{F_2+3F_3=F_2'\\ F_2'\leftrightarrow\frac{1}{7}F_3}}{\sim}\left[\begin{array}{ccc|c}1 & -4 & 3 & 0\\ 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 0\end{array}\right]\underset{F_1+4F_2}{\sim}\left[\begin{array}{ccc|c}1 & 0 & -1 & 0\\ 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 0\end{array}\right].$$

La solución general corresponde a $x = z$, $y = z$, luego el vector propio tiene la forma $\begin{bmatrix}1\\1\\1\end{bmatrix} z$. Cada vector de esta forma con $z \neq 0$ es un vector propio correspondiente a $\lambda_1 = 9$. Por lo que $\mathcal{B}_1 = \text{gen}\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$.

Para $\lambda_2 = 2$, se obtiene

$$(A - 2I)\vec{v} = \begin{bmatrix}2 & -1 & 6\\ 2 & -1 & 6\\ 2 & -1 & 6\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Para encontrar el vector propio correspondiente, se realizan operaciones por filas

$$\left[\begin{array}{ccc|c}2 & -1 & 6 & 0\\ 2 & -1 & 6 & 0\\ 2 & -1 & 6 & 0\end{array}\right]\underset{\substack{F_2-F_1\\ F_3-F_1}}{\sim}\left[\begin{array}{ccc|c}2 & -1 & 6 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{array}\right].$$

La solución general corresponde a $y = 2x + 6z$, luego el vector propio tiene la forma $\begin{bmatrix}1\\2\\0\end{bmatrix} x + \begin{bmatrix}0\\6\\1\end{bmatrix} z$. Cada vector de esta forma con $x, z \neq 0$ es un vector propio correspondiente a $\lambda_2 = 2$. Por lo que $\mathcal{B}_2 = \text{gen}\left\{\begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}0\\6\\1\end{bmatrix}\right\}$.

Teorema 2.8 *Sean $\lambda_1, \lambda_2, \ldots, \lambda_m$ valores propios distintos de la matriz A. Si para cada $k = 1, 2, \ldots, m$, S_k es un conjunto linealmente independiente de vectores propios de A correspondientes a λ_k, entonces:*

$$S = S_1 \cup S_2 \cup \ldots \cup S_m$$

es todavía un conjunto de vectores linealmente independiente.

Demostración.

Sea $S_k = \{\vec{x}_{k1}, \ldots, \vec{x}_{kr_k}\}$ un conjunto linealmente independiente de vectores propios de la matriz A, correspondiente al valor propio λ_k (para cada $k \in \{1, 2, \ldots, m\}$).

Para probar que $S = S_1 \cup S_2 \cup \ldots \cup S_m$ es linealmente independiente, consideremos una combinación lineal de los vectores en esa unión tal que

$$\underbrace{\left(a_{11}\vec{x}_{11} + \ldots + a_{1r_1}\vec{x}_{1r_1}\right)}_{\text{en } \mathcal{B}_1} + \ldots + \underbrace{\left(a_{m1}\vec{x}_{m1} + \ldots + a_{mr_m}\vec{x}_{mr_m}\right)}_{\text{en } \mathcal{B}_m} = \vec{0}.$$

Por el Teorema 2.7, se puede afirmar que cada suma entre paréntesis es $\vec{0}$. Como los vectores que participan en cada una de esas sumas son linealmente independientes, entonces se concluye que los coeficientes tienen que ser nulos, lo cual prueba que S es linealmente independiente. ■

La prueba del siguiente enunciado no es difícil si se explican algunos otros resultados. Su demostración se realiza en la siguiente sección.

Teorema 2.9 *Si λ_k es un valor propio de una matriz real A de tamaño $n \times n$, con multiplicidad algebraica r, entonces:*

$$1 \leq \textit{ multiplicidad geométrica de } \lambda_k \leq \textit{ multiplicidad algebraica de } \lambda_k.$$

Teorema 2.10 *Sea A una matriz no singular de tamaño $n \times n$, con valores propios no nulos $\lambda_1, \lambda_2, \ldots, \lambda_n$ y vectores propios $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. La matriz inversa A^{-1} tiene:*

i) Los valores propios de la forma $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$.

ii) Los mismos vectores propios de A.

Demostración.

i) Si los valores propios de A son diferentes de cero, entonces

$$\begin{aligned} p_A(\lambda) &= \det\big(A - \lambda I_n\big) = \det\big[A\big(I_n - \lambda A^{-1}\big)\big] \\ &= \det A \det\left[-\lambda\left(A^{-1} - \frac{1}{\lambda} I_n\right)\right] \end{aligned}$$

$$\begin{aligned} p_A(\lambda) &= \big(-\lambda\big)^n \det(A) \det\left(A^{-1} - \frac{1}{\lambda} I_n \right) \\ &= \big(-1\big)^n \lambda^n \det(A) p_{A^{-1}}\left(\frac{1}{\lambda}\right). \end{aligned}$$

Luego, se deduce que $1/\lambda$ es un valor propio de A^{-1} por cada valor λ de A.

ii) Si $A\vec{u} = \lambda\vec{u}$ entonces premultiplicando por A^{-1}, se tiene

$$\begin{aligned} A^{-1}\big(A\vec{u}\big) &= A^{-1}\big(\lambda\vec{u}\big) \\ \vec{u} &= \lambda A^{-1}\vec{u} \qquad \text{puesto que } \lambda \neq 0. \\ \frac{1}{\lambda}\vec{u} &= A^{-1}\vec{u}, \end{aligned}$$

por lo tanto, $\vec{u}$ es también un vector propio de A^{-1}. ■

2.1.1 Descomposición de Sylvester

En la siguiente sección se presenta una descomposición útil para matrices cuadradas; esta importante descomposición fue dada por Sylvester.

Definición 2.6 Matriz de proyección espectral

Sea A una matriz real de tamaño $n \times n$ que no tiene valores propios múltiples y sean $\vec{v}_k$, $\vec{w}_k$ los vectores propios a derecha e izquierda de A, es decir los vectores propios de A y A^t asociados al valor propio real λ_k. Se define la matriz de proyección espectral correspondiente a cada λ_k como:

$$\mathbf{E}\big(\lambda_k\big) = \frac{\vec{v}_k\vec{w}_k^t}{\langle\vec{w}_k, \vec{v}_k\rangle} = \frac{\vec{v}_k\vec{w}_k^t}{\vec{w}_k^t\vec{v}_k}. \tag{2.11}$$

Ejemplo 2.5 *Encuentre las matrices de proyección espectral de la siguiente matriz:*

$$A = \begin{bmatrix} 4 & -1 & 6 \\ 8 & 1 & 0 \\ 3 & -1 & 7 \end{bmatrix}.$$

Solución.

La ecuación característica asociada a la matriz A es

$$-\lambda^3 + 12\lambda^2 - 29\lambda + 18 = -\big(\lambda - 1\big)\big(\lambda - 2\big)\big(\lambda - 9\big) = 0.$$

De esta manera, los valores propios de A son $\lambda_1 = 1$, $\lambda_2 = 2$ y $\lambda_3 = 9$.

Para $\lambda_1 = 1$, se tiene

$$(A - I)\vec{v} \;=\; \begin{bmatrix} 3 & -1 & 6 \\ 8 & 0 & 0 \\ 3 & -1 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Para determinar el vector propio correspondiente, se realizan operaciones por filas

$$\left[\begin{array}{ccc|c} 3 & -1 & 6 & 0 \\ 8 & 0 & 0 & 0 \\ 3 & -1 & 6 & 0 \end{array}\right] \underset{\substack{F_3 - F_1 \\ F_2 \leftrightarrow \frac{1}{8}F_2}}{\sim} \left[\begin{array}{ccc|c} 3 & -1 & 6 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \underset{F_1 - 3F_2}{\sim} \left[\begin{array}{ccc|c} 0 & -1 & 6 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

La solución general corresponde a $x = 0$ y $y = 6z$, luego el vector propio tiene la forma $\begin{bmatrix} 0 \\ 6 \\ 1 \end{bmatrix} z$. Cada vector de esta forma con $z \neq 0$ es un vector propio correspondiente a $\lambda_1 = 1$. Por lo que $\mathcal{B}_1 = \text{gen}\left\{ \begin{bmatrix} 0 \\ 6 \\ 1 \end{bmatrix} \right\}$.

Para $\lambda_2 = 2$, se obtiene

$$(A - 2I)\vec{v} \;=\; \begin{bmatrix} 2 & -1 & 6 \\ 8 & -1 & 0 \\ 3 & -1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Para encontrar el vector propio correspondiente, se realizan operaciones por filas

$$\left[\begin{array}{ccc|c} 2 & -1 & 6 & 0 \\ 8 & -1 & 0 & 0 \\ 3 & -1 & 5 & 0 \end{array}\right] \underset{\substack{F_3 - F_1 \\ F_1 - F_2}}{\sim} \left[\begin{array}{ccc|c} -6 & 0 & 6 & 0 \\ 8 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \end{array}\right].$$

La solución general corresponde a $y = 8x$ y $z = x$, luego el vector propio tiene la forma $\begin{bmatrix} 1 \\ 8 \\ 1 \end{bmatrix} x$. Cada vector de esta forma con $x \neq 0$ es un vector propio correspondiente a $\lambda_2 = 2$. Por lo que $\mathcal{B}_2 = \text{gen}\left\{ \begin{bmatrix} 1 \\ 8 \\ 1 \end{bmatrix} \right\}$.

Para $\lambda_3 = 9$, se tiene

$$(A - 9I)\vec{v} = \begin{bmatrix} -5 & -1 & 6 \\ 8 & -8 & 0 \\ 3 & -1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Para encontrar el vector propio correspondiente, se realizan operaciones por filas

$$\left[\begin{array}{ccc|c} -5 & -1 & 6 & 0 \\ 8 & -8 & 0 & 0 \\ 3 & -1 & -2 & 0 \end{array}\right] \underset{\frac{1}{8}F_2}{\overset{}{\sim}} \left[\begin{array}{ccc|c} -5 & -1 & 6 & 0 \\ 1 & -1 & 0 & 0 \\ 8 & 0 & -8 & 0 \end{array}\right].$$

La solución general corresponde a $y = x$, $z = x$, luego el vector propio tiene la forma $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x$. Cada vector de esta forma con $x \neq 0$ es un vector propio correspondiente a $\lambda_3 = 9$. Por lo que $\mathcal{B}_3 = \text{gen}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Las matrices A y A^t tienen los mismos valores propios y realizando el mismo procedimiento descrito anteriormente, se tienen los siguientes vectores propios para la matriz A^t asociados a los valores propios $\lambda_1 = 1$, $\lambda_2 = 2$ y $\lambda_3 = 9$, respectivamente,

$$\vec{w}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \qquad \vec{w}_2 = \begin{bmatrix} 5 \\ 1 \\ -6 \end{bmatrix} \qquad \text{y} \qquad \vec{w}_3 = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}.$$

Luego, las matrices de proyección espectral $\mathbf{E}(\lambda_k)$ son

$$\mathbf{E}(1) = \begin{bmatrix} 0 & 0 & 0 \\ -6 & 0 & 6 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{E}(2) = \frac{1}{7}\begin{bmatrix} 5 & 1 & -6 \\ 40 & 8 & -48 \\ 5 & 1 & -6 \end{bmatrix} \quad \text{y} \quad \mathbf{E}(9) = \frac{1}{7}\begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}.$$

El lector puede verificar que $\sum_{i=1}^{3} \mathbf{E}(\lambda_i) = I_3$.

Teorema 2.11 *Sea A una matriz real de tamaño $n \times n$ que no tiene valores propios múltiples, entonces los vectores propios a derecha e izquierda asociados a valores propios reales distintos de A son ortogonales.*

Demostración.

Sean $\vec{v}_i$ y $\vec{w}_j$ los vectores propios a derecha e izquierda asociados a valores propios reales distintos λ_i y λ_j, respectivamente. Para comprobar que $\vec{v}_i \cdot \vec{w}_j = 0$, se calcula

$$\begin{aligned}
\lambda_i(\vec{v}_i \cdot \vec{w}_j) &= (\lambda_i \vec{v}_i)^t \vec{w}_j \\
&= (A\vec{v}_i)^t \vec{w}_j \qquad \text{puesto que } \vec{v}_i \text{ es un vector propio de } A \\
&= (\vec{v}_i^{\,t}\, A^t)\vec{w}_j = \vec{v}_i^{\,t}(A^t \vec{w}_j) \text{ reagrupando términos} \\
&= \vec{v}_i^{\,t}(\lambda_j \vec{w}_j) \qquad \text{ya que } \vec{w}_j \text{ es un vector propio de } A^t \\
&= \lambda_j \vec{v}_i^{\,t} \vec{w}_j = \lambda_j(\vec{v}_i \cdot \vec{w}_j).
\end{aligned}$$

Luego $(\lambda_i - \lambda_j)\vec{v}_i \cdot \vec{w}_j = 0$ y como $\lambda_i - \lambda_j \neq 0$, entonces $\vec{v}_i \cdot \vec{w}_j = 0$. ■

Teorema 2.12 *Sea A una matriz real de tamaño $n \times n$ con valores propios reales distintos $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces las matrices de proyección espectral $\mathbf{E}(\lambda_k)$ definidas en* (2.11) *satisfacen las siguientes propiedades:*

a) $\mathbf{E}(\lambda_i)\mathbf{E}(\lambda_j) = \begin{cases} \mathbf{E}(\lambda_i) & \text{si } i = j, \\ \mathrm{O} & \text{si } i \neq j. \end{cases}$ *b)* $\displaystyle\sum_{k=1}^{n} \mathbf{E}(\lambda_k) = I_n,$

c) Cada $\mathbf{E}(\lambda_k)$ conmuta con A, es decir $A\mathbf{E}(\lambda_k) = \mathbf{E}(\lambda_k)A$.

Demostración.

a) Sean $\vec{v}_i$ y $\vec{w}_j$ los vectores propios a derecha e izquierda asociados a valores propios reales distintos λ_i y λ_j, respectivamente. Por el Teorema 2.11 estos vectores son ortogonales, por lo tanto de la definición de $\mathbf{E}(\lambda_k)$ dada en (2.11), se tiene

$$\mathbf{E}(\lambda_i)\mathbf{E}(\lambda_j) = \frac{\vec{v}_i \vec{w}_i^{\,t}}{\vec{w}_i^{\,t}\vec{v}_i} \frac{\vec{v}_j \vec{w}_j^{\,t}}{\vec{w}_j^{\,t}\vec{v}_j} = \frac{\vec{v}_i}{\vec{w}_i^{\,t}\vec{v}_i} \vec{w}_i^{\,t}\vec{v}_j \frac{\vec{w}_j^{\,t}}{\vec{w}_j^{\,t}\vec{v}_j} = \mathrm{O},$$

de manera análoga

$$\mathbf{E}(\lambda_i)\mathbf{E}(\lambda_i) = \frac{\vec{v}_i \vec{w}_i^{\,t}}{\vec{w}_i^{\,t}\vec{v}_i} \frac{\vec{v}_i \vec{w}_i^{\,t}}{\vec{w}_i^{\,t}\vec{v}_i} = \vec{v}_i \frac{\vec{w}_i^{\,t}\vec{v}_i}{\vec{w}_i^{\,t}\vec{v}_i} \frac{\vec{w}_i^{\,t}}{\vec{w}_i^{\,t}\vec{v}_i} = \mathbf{E}(\lambda_i).$$

b) Queda como ejercicio para el lector.

c) Al premultiplicar por A cualquier matriz $\mathbf{E}(\lambda_k)$, se obtiene

$$A\mathbf{E}(\lambda_k) = A\vec{v}_k \frac{\vec{w}_k^t}{\vec{v}_k^t \vec{w}_k} = \lambda_k \frac{\vec{v}_k \vec{w}_k^t}{\vec{v}_k^t \vec{w}_k} = \lambda_k \mathbf{E}(\lambda_k), \tag{2.12}$$

y al multiplicar por A cualquier matriz $\mathbf{E}(\lambda_k)$, se llega a

$$\mathbf{E}(\lambda_k)A = \frac{\vec{v}_k}{\vec{v}_k^t \vec{w}_k} (A^t \vec{w}_k)^t = \lambda_k \frac{\vec{v}_k \vec{w}_k^t}{\vec{v}_k^t \vec{w}_k} = \lambda_k \mathbf{E}(\lambda_k). \tag{2.13}$$

Luego, $A\mathbf{E}(\lambda_k) = \mathbf{E}(\lambda_k)A$ para $k = 1, 2, \ldots, n$. ■

Teorema 2.13 Descomposición de Sylvester

Sea A una matriz real de tamaño $n \times n$ con valores propios reales distintos $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces A se puede escribir como

$$A = \sum_{k=1}^{n} \lambda_k \mathbf{E}(\lambda_k), \tag{2.14}$$

donde la matriz $\mathbf{E}(\lambda_k)$ es dada en (2.11).

Demostración.

Puesto que cada $\mathbf{E}(\lambda_k)$ conmuta con A, al sumar las expresiones obtenidas en (2.12) y (2.13), se tiene

$$\underbrace{\sum_{k=1}^{n} A\mathbf{E}(\lambda_k)}_{} = \underbrace{\sum_{k=1}^{n} \mathbf{E}(\lambda_k)A}_{} = \sum_{k=1}^{n} \lambda_k \mathbf{E}(\lambda_k)$$

$$A \sum_{k=1}^{n} \mathbf{E}(\lambda_k) = \left[\sum_{k=1}^{n} \mathbf{E}(\lambda_k) \right] A = \sum_{k=1}^{n} \lambda_k \mathbf{E}(\lambda_k),$$

empleando la propiedad *b*) del Teorema 2.12 se completa la prueba. ■

En virtud del Teorema 2.13, se tiene que:

1. La descomposición de Sylvester de la matriz simétrica A^tA es

$$A^tA = \sum_{i=1}^{n} \lambda_i^2 \|\vec{v}_i\|^2 \frac{\vec{w}_i\vec{w}_i^t}{\left(\vec{w}_i^t\vec{v}_i\right)^2} + \sum_{i=1}^{m}\sum_{j>i}^{m} \lambda_i\lambda_j \frac{\|\vec{v}_i\|\,\|\vec{v}_j\|\,\vec{w}_i\vec{w}_j^t}{\left(\vec{w}_i^t\vec{v}_i\right)\left(\vec{w}_j^t\vec{v}_j\right)} \cos\theta_{ij},$$

donde $\theta_{ij} = \sphericalangle(\vec{v}_i, \vec{v}_j)$.

2. La descomposición de Sylvester de la matriz simétrica AA^t es

$$AA^t = \sum_{i=1}^{n} \lambda_i^2 \|\vec{w}_i\|^2 \frac{\vec{v}_i\vec{v}_i^t}{\left(\vec{v}_i^t\vec{w}_i\right)^2} + \sum_{i=1}^{m}\sum_{j>i}^{m} \lambda_i\lambda_j \frac{\|\vec{w}_i\|\,\|\vec{w}_j\|\,\vec{v}_i\vec{v}_j^t}{\left(\vec{v}_i^t\vec{w}_i\right)\left(\vec{v}_j^t\vec{w}_j\right)} \cos\varphi_{ij},$$

donde $\varphi_{ij} = \sphericalangle(\vec{w}_i, \vec{w}_j)$.

Ejemplo 2.6 *Encuentre la descomposición de Sylvester para la matriz dada en el Ejemplo 2.5.*

Solución.

En el Ejemplo 2.5 se obtuvo que los valores propios de A eran $\lambda_1 = 1$, $\lambda_2 = 2$ y $\lambda_3 = 9$, y las matrices $\mathbf{E}(\lambda_k)$ fueron

$$\mathbf{E}(1) = \begin{bmatrix} 0 & 0 & 0 \\ -6 & 0 & 6 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{E}(2) = \frac{1}{7}\begin{bmatrix} 5 & 1 & -6 \\ 40 & 8 & -48 \\ 5 & 1 & -6 \end{bmatrix} \quad \text{y} \quad \mathbf{E}(9) = \frac{1}{7}\begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}.$$

El lector puede verificar que $A = \sum_{i=1}^{3} \lambda_i \mathbf{E}(\lambda_i)$.

Corolario 2.13.1 *Si A es una matriz real de tamaño $n \times n$ con valores propios reales λ_k de multiplicidad algebraica r_k y multiplicidad geométrica igual a r_k, entonces A se puede escribir como*

$$A = \sum_{k} \lambda_k \mathbf{E}^*(\lambda_k), \tag{2.15}$$

donde las matrices $\mathbf{E}^*(\lambda_k)$ vienen dadas por:

$$\mathbf{E}^*(\lambda_k) = P(\lambda_k)\left\{\left[P^*(\lambda_k)\right]^t P(\lambda_k)\right\}^{-1}\left[P^*(\lambda_k)\right]^t, \tag{2.16}$$

con $P(\lambda_k) = [\vec{v}_{k1} \quad \vec{v}_{k2} \quad \dots \quad \vec{v}_{kr}]$ *y* $P^*(\lambda_k) = [\vec{w}_{k1} \quad \vec{w}_{k2} \quad \dots \quad \vec{w}_{kr}]$ *matrices de tamaño* $n \times r_k$ *cuyas respectivas columnas son los vectores propios de* A *y* A^t *asociados a* λ_k.

Demostración.

Queda como ejercicio para el lector. ■

Ejemplo 2.7 *Encuentre la descomposición de Sylvester para la matriz dada en el Ejemplo 2.3.*

Solución.

En el Ejemplo 2.4, se obtuvo que los valores propios asociados a la matriz A eran $\lambda_1 = 9$ y $\lambda_2 = \lambda_3 = 2$, y los respectivos vectores propios de A eran

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} 0 \\ 6 \\ 1 \end{bmatrix}.$$

Por otra parte, los vectores propios de A^t son

$$\vec{w}_1 = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}, \qquad \vec{w}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{w}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}.$$

Luego la matriz $P(2) = \begin{bmatrix} \vec{v}_2 & \vec{v}_3 \end{bmatrix}$ y $P^*(2) = \begin{bmatrix} \vec{w}_2 & \vec{w}_3 \end{bmatrix}$, por lo tanto

$$\left[P^*(2)\right]^t P(2) = -\begin{bmatrix} 1 & 6 \\ 2 & 5 \end{bmatrix}, \qquad \left\{\left[P^*(2)\right]^t P(2)\right\}^{-1} = \frac{1}{7}\begin{bmatrix} 5 & -6 \\ -2 & 1 \end{bmatrix},$$

entonces

$$\mathbf{E}(9) = \frac{1}{7}\begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}, \qquad \mathbf{E}^*(2) = \frac{1}{7}\begin{bmatrix} 5 & 1 & -6 \\ -2 & 8 & -6 \\ -2 & 1 & 1 \end{bmatrix}.$$

El lector puede verificar que

$$\mathbf{E}(9) + \mathbf{E}^*(2) = I_3 \qquad \text{y} \qquad 9\mathbf{E}(9) + 2\mathbf{E}^*(2) = A.$$

Ejercicios 2.1

1. Para cada una de las siguientes matrices:

$$a.\begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}. \quad b.\begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}. \quad c.\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}. \quad d.\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}.$$

$$e.\begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}. \quad f.\begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}. \quad g.\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}. \quad h.\begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$$

   i) Calcule los valores propios y los espacios propios.

   ii) Determine las respectivas proyecciones espectrales.

   iii) Encuentre la correspondiente descomposición de Sylvester.

2. Sea la matriz $A = \begin{bmatrix} 1 & k \\ 1 & 1 \end{bmatrix}$, con k una constante arbitraria. ¿Para qué valores de k la matriz A tiene dos valores propios reales distintos?

3. Si A es una matriz diagonal de tamaño $n \times n$, muestre que sus valores propios son las entradas de su diagonal.

4. Si A es una matriz triangular de tamaño $n \times n$, muestre que sus valores propios son las entradas de su diagonal principal.

5. Si A es una matriz real de tamaño $n \times n$, muestre que es invertible si y solo si el número 0 no es un valor propio de A.

6. Si A es una matriz real de tamaño $n \times n$ con la propiedad de que la suma de los elementos de sus filas es siempre igual a un número s, muestre que s es un valor propio de A.

7. Sea A una matriz de tamaño $n \times n$ y $\lambda_1, \lambda_2, \ldots, \lambda_n$ sus valores propios, demuestre que:

   a) La matriz A^t tiene los mismos valores propios.

   b) La matriz kA tiene valores propios $k\lambda_1, k\lambda_2, \ldots, k\lambda_n$.

   c) La matriz A^k (donde k es un entero positivo) tiene valores propios $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$.

2.2 Matrices semejantes y diagonalización

En la sección anterior se desarrolló parte de la terminología y de las propiedades de los valores propios y vectores propios. En este apartado continuaremos estudiando los valores propios, debido a que estos números son cruciales en muchas consideraciones, incluyendo las representaciones de matrices en ciertas formas que permiten trabajar la solución de problemas de manera más fácil.

Definición 2.7 Matrices congruentes

Dos matrices reales A y B de tamaño $n \times n$ son congruentes si existe una matriz P no singular de componentes reales de tamaño $n \times n$ tal que

$$A = P^t B P. \tag{2.17}$$

Ejemplo 2.8 *Determine si las siguientes matrices son congruentes:*

$$A = \begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} \qquad y \qquad B = \begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix}.$$

Solución.

Veamos si existe una matriz P tal que $B = P^tAP$. En particular, P puede ser una matriz triangular superior o en otras palabras,

$$\begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}^t \begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$$
$$= \begin{bmatrix} a^2 & a\big(b+4d\big) \\ a\big(b+4d\big) & b^2+8bd+d^2 \end{bmatrix}.$$

En consecuencia,

$$a^2 = 1, \qquad a(b+4d) = 1 \qquad \text{y} \qquad b^2+8bd+d^2 = -14.$$

Si se despeja b de la segunda ecuación y se reemplaza en la tercera, se tiene que

$$\left(\frac{1}{a}-4d\right)^2 + 8d\left(\frac{1}{a}-4d\right) + d^2 = \frac{1}{a^2} - 15d^2$$
$$1 - 15d^2 = -14.$$

Luego $d^2 = 1$, por lo tanto la matriz P puede ser

$$\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \qquad \text{ó} \qquad \begin{bmatrix} -1 & -5 \\ 0 & 1 \end{bmatrix} \qquad \text{ó} \qquad \begin{bmatrix} 1 & 5 \\ 0 & -1 \end{bmatrix}.$$

Teorema 2.14 *Dos matrices reales A y B de tamaño $n \times n$ son congruentes si y solo si tienen el mismo rango.*

Demostración.

Como A y B son matrices congruentes de tamaño $n \times n$, existe una matriz no singular P tal que $B = P^tAP$, entonces:

$$\begin{aligned} \rho(B) &= \rho\big[\big(P^tA\big)P\big] = \rho\big(P^tA\big) && \text{puesto que } P \text{ es no singular,} \\ &= \rho\big(A\big) && \text{por ser } P^t \text{ también no singular.} \end{aligned}$$

Aquí se utilizó la propiedad (x) dada en el Teorema 1.37. ■

Teorema 2.15 *La congruencia de matrices de tamaño $n \times n$ cumple las propiedades de relación de equivalencia, es decir es*

a) Reflexiva: A es congruente a A.

b) Simétrica: si A es congruente a B, entonces B es congruente a A.

c) Transitiva: si A es congruente a B y B es congruente a C, entonces A es congruente a C.

Demostración.
Queda como ejercicio para el lector. ■

Para las matrices cuadradas, además del concepto de congruencia, se tiene otro de mayor utilidad o generalidad: el de *similaridad*.

Definición 2.8 Matrices semejantes

Una matriz A de tamaño $n \times n$ es semejante (o similar) a una matriz B de tamaño $n \times n$ si existe una matriz no singular P de tamaño $n \times n$ tal que

$$B = P^{-1}AP. \tag{2.18}$$

De manera análoga, se dice que A y B son semejantes si y solo si existe una matriz no singular P tal que

$$PB = AP. \tag{2.19}$$

Ejemplo 2.9 *Determine si A y B son semejantes, dado que*

$$A = \begin{bmatrix} -1 & 0 \\ 1 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} -3 & 5 \\ -2 & 4 \end{bmatrix} \qquad y \qquad P = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

Solución.
Se realizan los productos AP y PB:

$$AP = \begin{bmatrix} -1 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & 3 \end{bmatrix}.$$

$$PB = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -3 & 5 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & 3 \end{bmatrix}.$$

Así, $AP = PB$. Como $\det(P) = 1 \neq 0$, entonces P es no singular. Y por la ecuación (2.19), se tiene que A y B son semejantes.

Teorema 2.16 *Las matrices semejantes tienen el mismo polinomio característico y, por tanto, los mismos valores propios.*

Demostración.

Como A y B son matrices semejantes de tamaño $n \times n$, $B = P^{-1}AP$. Entonces

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}[AP - \lambda P] = P^{-1}[A - \lambda I]P.$$

Por consiguiente,

$$\begin{aligned} \det(B - \lambda I) &= \det\big[P^{-1}(A - \lambda I)P\big] = \det\big(P^{-1}\big)\det(A - \lambda I)\det(P) \\ &= \det\big(P^{-1}\big)\det(P)\det(A - \lambda I) = \det(A - \lambda I). \end{aligned}$$

Esto significa que A y B tienen la misma ecuación característica, y como los valores propios son raíces de la ecuación característica, entonces tienen también los mismos valores propios. ■

Ejemplo 2.10 *Para las matrices A y B dadas en el Ejemplo 2.9, muestre que tienen el mismo polinomio característico.*

Solución.

Tenemos que

$$\det(A - \lambda I) = \begin{vmatrix} -1 - \lambda & 0 \\ 1 & 2 - \lambda \end{vmatrix} = \lambda^2 - \lambda - 2,$$

y

$$\det(B - \lambda I) = \begin{vmatrix} -3 - \lambda & 5 \\ -2 & 4 - \lambda \end{vmatrix} = \lambda^2 - \lambda - 2.$$

Como $\det(A - \lambda I) = \det(B - \lambda I)$, las matrices A y B tienen el mismo polinomio característico, y por lo tanto los mismos valores propios.

Teorema 2.17 *La semejanza de matrices de tamaño $n \times n$ cumple las propiedades de relación de equivalencia, es decir es*

a) Reflexiva: A es semejante a A.

b) Simétrica: si A es semejante a B, entonces B es semejante a A.

c) Transitiva: si A es semejante a B y B es semejante a C, entonces A es semejante a C.

Demostración.
Queda como ejercicio para el lector. ■

Teorema 2.18 *Si B es una matriz semejante a A con $B = P^{-1}AP$, entonces $\vec{v}$ es un vector propio de A asociado con el valor propio λ si y solo si $P^{-1}\vec{v}$ es un vector propio de B asociado con el valor propio λ.*

Demostración.
Si $\vec{v}$ es un vector propio de A, se tiene que

$$\begin{aligned} A\vec{v} &= \lambda\vec{v} \\ \big(PBP^{-1}\big)\vec{v} &= \lambda\vec{v} && \text{puesto que } B \text{ es semejante a } A, \\ B\big(P^{-1}\vec{v}\big) &= \lambda P^{-1}\vec{v} && \text{puesto que } P \text{ es no singular,} \end{aligned}$$

lo cual completa la prueba. ■

Ejemplo 2.11 *Para cada una de las matrices A y B dadas en el Ejemplo 2.9, determine sus vectores propios.*

Solución.
Para la matriz A, se tiene que los valores propios son $\lambda_1 = -1$ y $\lambda_2 = 2$ y sus correspondientes vectores propios son $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ y $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, respectivamente.

Para la matriz B, se tiene que los valores propios son $\lambda_1 = -1$ y $\lambda_2 = 2$ y sus correspondientes vectores propios son $\begin{bmatrix} -5 \\ -2 \end{bmatrix}$ y $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, respectivamente.

El lector puede verificar que los vectores propios de B son iguales a los vectores propios de A premultiplicados por la inversa de la matriz P dada en el Ejemplo 2.9.

Con lo que se ha estudiado hasta ahora en esta sección, se puede presentar una demostración del Teorema 2.9.

Demostración. Teorema 2.9

Sea $\mathcal{B}_k = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ base del espacio propio correspondiente al valor propio λ_k, donde m es la multiplicidad geométrica de λ_k. Se extiende $\mathcal{B}_k$ hasta completar una base de $\mathbb{R}^n$, digamos

$$\mathfrak{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m, \vec{v}_{m+1}, \vec{v}_{m+2}, \ldots, \vec{v}_n\}.$$

En esta base, la matriz A está particionada como

$$[A]_{\mathfrak{B}} = \begin{bmatrix} \lambda_k I_m & A_{12} \\ 0 & A_{22} \end{bmatrix}.$$

Luego A y $[A]_{\mathfrak{B}}$ son matrices semejantes, es decir tienen el mismo polinomio característico y los mismos valores propios con idénticas multiplicidades algebraicas. Así que el polinomio característico de A es

$$p_A(\lambda) = p_{[A]_{\mathfrak{B}}}(\lambda) = (\lambda_k - \lambda)^m p_D(\lambda),$$

por lo tanto λ_k aparece como raíz de $p_A(\lambda)$ por lo menos m veces y, por consiguiente, la multiplicidad algebraica de λ_k es mayor o igual a m. ■

Definición 2.9 Matriz diagonalizable

Una matriz A de tamaño $n \times n$ es diagonalizable si existe una matriz diagonal D tal que A es semejante a D.

Este resultado es muy importante ya que las matrices diagonales poseen muchas propiedades que permiten trabajar fácilmente con ellas, véase el Teorema 1.14.

Teorema 2.19 *Una matriz A de tamaño $n \times n$ es diagonalizable si y solo si A tiene n vectores propios linealmente independientes. En tal caso, si $A = PDP^{-1}$ donde D es diagonal, entonces los elementos de la diagonal de D son los valores propios de A y las columnas de P son los vectores propios correspondientes.*

Demostración.

Primero se supone que A es diagonalizable. Entonces existe una matriz no singular P tal que $P^{-1}AP = D$ es diagonal. Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ los elementos de la diagonal principal de D y sean $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ los vectores columna de la matriz P, entonces

$$\begin{aligned} PD &= \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \ldots & \lambda_n\vec{v}_n \end{bmatrix}. \end{aligned}$$

Pero como $AP = \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}$ y por otra parte $P^{-1}AP = D$, se tiene que $AP = PD$, lo cual implica

$$\begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \ldots & \lambda_n\vec{v}_n \end{bmatrix}.$$

En otras palabras, $A\vec{v}_i = \lambda_i\vec{v}_i$ para todo vector columna $\vec{v}_i$. Esto significa que los vectores columna $\vec{v}_i$ de P son vectores propios de A. Además, como P es una matriz no singular, entonces sus vectores columna son linealmente independientes. Así, A tiene n vectores propios linealmente independientes.

Recíprocamente, suponga que A tiene n vectores propios linealmente independientes $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ con valores propios asociados $\lambda_1, \lambda_2, \ldots, \lambda_n$. Sea P la matriz cuyas columnas son estos n vectores propios, es decir, $P = [\vec{v}_1 \;\; \vec{v}_2 \;\; \ldots \;\; \vec{v}_n]$. Como todo $\vec{v}_i$ es un vector propio de A, entonces se tiene que $A\vec{v}_i = \lambda_i\vec{v}_i$ y

$$AP = A\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\vec{v}_1 & \lambda_2\vec{v}_2 & \ldots & \lambda_n\vec{v}_n \end{bmatrix}.$$

Nótese que la matriz del lado derecho de esta ecuación puede escribirse como el siguiente producto de matrices

$$AP = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} = PD.$$

Por último, como los vectores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ son linealmente independientes, entonces P es no singular y se puede escribir la ecuación $AP = PD$ como $P^{-1}AP = D$, lo cual significa que A es diagonalizable. ■

Corolario 2.19.1 *Si una matriz A de tamaño $n \times n$ tiene n valores propios distintos, entonces A es diagonalizable.*

Demostración.

Sean $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ los vectores propios correspondientes a los n valores propios distintos de la matriz A. Entonces por el Teorema 2.6, se tiene que el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ es linealmente independiente. Luego por el Teorema 2.19, A es diagonalizable. ■

A continuación, se presenta un procedimiento para diagonalizar una matriz A de tamaño $n \times n$.

Procedimiento para diagonalizar una matriz cuadrada

Sea A una matriz real de tamaño $n \times n$.

i) Determine n vectores propios $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ de A, con valores propios correspondientes $\lambda_1, \lambda_2, \ldots, \lambda_n$. Si no existen n vectores propios linealmente independientes, entonces A no es diagonalizable.

ii) Obtenga P como la matriz cuyas columnas son los vectores propios obtenidos en el paso i. Es decir,

$$P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix}.$$

iii) La matriz diagonal $D = P^{-1}AP$ tendrá los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$ en su diagonal principal (y ceros en el resto). La ubicación de los vectores propios en la matriz P determina la posición en que aparecen los valores propios sobre la diagonal de D.

Ejemplo 2.12 *Determine si la matriz dada a continuación es diagonalizable:*

$$A = \begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}.$$

Solución.

La ecuación característica asociada a la matriz A es

$$-\lambda^3 + 4\lambda^2 - 4\lambda = -\lambda(\lambda - 2)^2 = 0.$$

Luego, los valores propios son $\lambda_1 = 0$ y $\lambda_2 = 2$ (de multiplicidad algebraica 2). El vector propio correspondiente a $\lambda_1 = 0$ es $\vec{v}_1 = \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$ y los correspondientes a $\lambda_2 = 2$ son $\vec{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$. Entonces

$$P = \begin{bmatrix} -1 & -2 & 3 \\ 3 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad \text{y} \qquad P^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix},$$

luego

$$\begin{aligned} P^{-1}AP &= \frac{1}{2}\begin{bmatrix} 1 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}\begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}\begin{bmatrix} -1 & -2 & 3 \\ 3 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}\begin{bmatrix} 0 & -4 & 6 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}. \end{aligned}$$

Por lo tanto, A es diagonalizable.

Ejemplo 2.13 Una matriz no diagonalizable

Determine si la siguiente matriz es diagonalizable:

$$A = \begin{bmatrix} 3 & -2 \\ 8 & -5 \end{bmatrix}.$$

Solución.

La ecuación característica de A es $\lambda^2 + 2\lambda + 1 = (\lambda + 1)^2 = 0$, luego $\lambda = -1$ es un valor propio de multiplicidad algebraica 2. Entonces,

$$(A - \lambda I)\vec{v} = (A + I)\vec{v} = \begin{bmatrix} 4 & -2 \\ 8 & -4 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Esto lleva al vector propio $\vec{v}_1 = \begin{bmatrix}1\\2\end{bmatrix}$. Por lo tanto, A no contiene dos vectores propios linealmente independientes, entonces se concluye que la matriz A no es diagonalizable.

Teorema 2.20 *Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ valores propios distintos de una matriz A de tamaño $n \times n$, entonces:*

$$\mathrm{tr}_1(A) = \sum_{i=1}^{n} \lambda_i \qquad y \qquad \det(A) = \prod_{i=1}^{n} \lambda_i.$$

Demostración.

Como A es diagonalizable, entonces $A = PDP^{-1}$, luego

$$\mathrm{tr}_1(A) = \mathrm{tr}_1[P(DP^{-1})] = \mathrm{tr}_1[(DP^{-1})P] = \mathrm{tr}_1(D) = \sum_{i=1}^{n} \lambda_i.$$

Por otra parte,

$$\det(A) = |A| = |P(DP^{-1})| = |P||DP^{-1}| = |D||P^{-1}||P| = |D| = \prod_{i=1}^{n} \lambda_i.$$

■

Ejercicios 2.2

1. Para las siguientes matrices, determine (en caso de ser posible) una matriz P no singular tal que $P^{-1}AP$ sea diagonal

$$a.\begin{bmatrix}1 & 1\\ -1 & 3\end{bmatrix}. \qquad b.\begin{bmatrix}1 & 1\\ 3 & -1\end{bmatrix}. \qquad c.\begin{bmatrix}1 & -1\\ -1 & 3\end{bmatrix}. \qquad d.\begin{bmatrix}1 & 2\\ 4 & 3\end{bmatrix}.$$

$$e.\begin{bmatrix}2 & 2 & 1\\ 1 & 3 & 1\\ 1 & 2 & 2\end{bmatrix}. \quad f.\begin{bmatrix}3 & -1 & 4\\ -4 & 1 & -5\\ 5 & 3 & 2\end{bmatrix}. \quad g.\begin{bmatrix}3 & 1 & -1\\ 2 & 2 & -1\\ 2 & 2 & 0\end{bmatrix}. \quad h.\begin{bmatrix}1 & 3 & -2\\ 3 & 1 & 2\\ -1 & 1 & 1\end{bmatrix}.$$

2. Muestre que las trazas de matrices semejantes son iguales.

3. Si A y B son semejantes, demuestre que tienen el mismo determinante.

4. Sean A una matriz diagonalizable de tamaño $n \times n$ y P una matriz no singular de tamaño $n \times n$ tales que $B = P^{-1}AP$ sea la forma diagonal de A. Pruebe que

   a) $B^k = P^{-1}A^kP$, donde k es un entero positivo.

   b) $A^k = PB^kP^{-1}$, donde k es un entero positivo.

5. Sea $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Con elementos reales, demuestre que A es diagonalizable si $-4bc < (a-d)^2$ y no diagonalizable si $-4bc > (a-d)^2$.

2.3 Valores propios complejos

Puesto que la ecuación característica de una matriz real de tamaño $n \times n$ es un polinomio de grado n, por el *Teorema Fundamental del Álgebra* se sabe que cualquier polinomio de grado n con coeficientes reales (o complejos) tiene exactamente n raíces (contando multiplicidades). En las secciones anteriores, desarrollamos la teoría para valores propios y vectores propios reales. En esta sección estudiaremos los valores propios y vectores propios complejos.

Definición 2.10 *Sea A una matriz real de tamaño $n \times n$. El número complejo λ es un valor propio de A si existe un vector no nulo $\vec{v} \in \mathbb{C}^n$ tal que*

$$A\vec{v} = \lambda\vec{v}. \tag{2.20}$$

Todo vector $\vec{v}$ no nulo que satisfaga (2.20) *es un vector propio de A asociado al valor propio λ.*

Ejemplo 2.14 *Sea* $A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$. *Determine los valores propios y vectores propios de A.*

Solución.

Primero se calcula el determinante de la matriz $A - \lambda I$:

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 1-\lambda & -1 \\ 2 & 1-\lambda \end{vmatrix} \\ &= (1-\lambda)^2 + 2 = \big[(1-\lambda) + i\sqrt{2}\big]\big[(1-\lambda) - i\sqrt{2}\big]. \end{aligned}$$

De esta manera, los valores propios de A son los complejos, a saber

$$\lambda_1 = 1 + i\sqrt{2} \qquad \text{y} \qquad \lambda_2 = 1 - i\sqrt{2}.$$

Para $\lambda_1 = 1 + i\sqrt{2}$, se tiene

$$\big[A - \big(1 + i\sqrt{2}\big)I\big]\vec{v} = \begin{bmatrix} -i\sqrt{2} & -1 \\ 2 & -i\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Para encontrar el vector propio correspondiente, se realizan operaciones por filas:

$$\left[\begin{array}{cc|c} -i\sqrt{2} & -1 & 0 \\ 2 & -i\sqrt{2} & 0 \end{array}\right] \underset{i\sqrt{2}F_1}{\sim} \left[\begin{array}{cc|c} 2 & -i\sqrt{2} & 0 \\ 2 & -i\sqrt{2} & 0 \end{array}\right] \underset{F_2 - F_1}{\sim} \left[\begin{array}{cc|c} 2 & -i\sqrt{2} & 0 \\ 0 & 0 & 0 \end{array}\right].$$

La solución general corresponde a $2x = i\sqrt{2}y$, luego el vector propio tiene la forma $\begin{bmatrix} 1 \\ -i\sqrt{2} \end{bmatrix} x$. Cada vector de esta forma con $x \neq 0$ es un vector propio correspondiente a $\lambda_1 = 1 + i\sqrt{2}$. Por lo tanto $\mathcal{B}_1 = \text{gen}\left\{ \begin{bmatrix} 1 \\ -i\sqrt{2} \end{bmatrix} \right\}$.

Para $\lambda_2 = 1 - i\sqrt{2}$, se tiene

$$\big[A - \big(1 - i\sqrt{2}\big)I\big]\vec{v} = \begin{bmatrix} i\sqrt{2} & -1 \\ 2 & i\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Para encontrar el vector propio correspondiente, se realizan operaciones por filas

$$\begin{bmatrix} i\sqrt{2} & -1 & | & 0 \\ 2 & i\sqrt{2} & | & 0 \end{bmatrix} \underset{-i\sqrt{2}F_1}{\sim} \begin{bmatrix} 2 & i\sqrt{2} & | & 0 \\ 2 & i\sqrt{2} & | & 0 \end{bmatrix} \underset{F_2 - F_1}{\sim} \begin{bmatrix} 2 & i\sqrt{2} & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}.$$

La solución general corresponde a $2x = -i\sqrt{2}y$, luego el vector propio tiene la forma $\begin{bmatrix} 1 \\ i\sqrt{2} \end{bmatrix} x$. Cada vector de esta forma con $x \neq 0$ es un vector propio correspondiente a $\lambda_2 = 1 - i\sqrt{2}$. Por lo tanto, $\mathcal{B}_2 = \text{gen}\left\{ \begin{bmatrix} 1 \\ i\sqrt{2} \end{bmatrix} \right\}$.

Teorema 2.21 *Sea A una matriz real de tamaño $n \times n$, entonces:*

i) Los valores propios de A cuando son complejos ocurren en pares conjugados.

ii) Los vectores propios correspondientes a valores propios complejos, son conjugados complejos entre sí.

Demostración.

i) Si A es una matriz real de tamaño $n \times n$, su polinomio característico se puede reescribir como

$$p_A(\lambda) = \det(A - \lambda I_n) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0,$$

donde cada $c_i \in \mathbb{R}$. Por lo tanto,

$$\overline{p_A(\lambda)} = \overline{\det(A - \lambda I_n)} = \det\big(\overline{A - \lambda I_n}\big) = \det\big(A - \overline{\lambda} I_n\big) = p_A\big(\overline{\lambda}\big).$$

Si λ_0 es una raíz de $p_A(\lambda)$, entonces

$$\overline{p_A(\lambda_0)} = p_A\big(\overline{\lambda_0}\big) = 0.$$

En consecuencia, $\overline{\lambda_0}$ es también un valor propio de A.

ii) Si λ es un valor propio complejo de A con un vector propio correspondiente $\vec{u} \in \mathbb{C}^n$, entonces en virtud del Teorema 1.42, se tiene

$$\begin{aligned} A\vec{u} &= \lambda\vec{u} \\ \overline{A\vec{u}} &= \overline{\lambda\vec{u}} \\ A\,\overline{\vec{u}} &= \overline{\lambda}\,\overline{\vec{u}}. \end{aligned}$$

Luego, $\overline{\vec{u}}$ es también vector propio de A, pero asociado al valor propio $\overline{\lambda}$. ■

Ejemplo 2.15 *Considere la matriz A del Ejemplo 2.14 y aplique el Teorema 2.21.*

Solución.

Para la matriz A cuyas componentes son reales, se obtuvo un valor propio $\lambda_1 = 1 + i\sqrt{2}$ con vector propio asociado $\vec{u}_1 = \begin{bmatrix} 1 \\ -i\sqrt{2} \end{bmatrix}$, y para el otro valor propio $\lambda_2 = 1 - i\sqrt{2}$, un vector propio asociado $\vec{u}_2 = \begin{bmatrix} 1 \\ i\sqrt{2} \end{bmatrix}$. Claramente, se nota que $\lambda_2 = \overline{\lambda}_1$ y que $\vec{u}_2 = \overline{\vec{u}}_1$.

Teorema 2.22 *Sea A una matriz real de tamaño* 2×2 *con un valor propio complejo* $\lambda = a + bi$ $(b \neq 0)$ *y vector propio correspondiente* $\vec{u} \in \mathbb{C}^2$. *Entonces*

$$\begin{aligned} ARe(\vec{u}) &= aRe(\vec{u}) - bIm(\vec{u}) \\ AIm(\vec{u}) &= bRe(\vec{u}) + aIm(\vec{u}), \end{aligned} \tag{2.21}$$

además, $Re(\vec{u})$ *y* $Im(\vec{u})$ *son vectores linealmente independientes.*

Demostración.

Sea $\vec{u} \in \mathbb{C}^2$ un vector propio de A, por lo tanto

$$\begin{aligned} A\vec{u} &= \lambda\vec{u} \\ A\big[Re(\vec{u}) + i\ Im(\vec{u})\big] &= (a+bi)\big[Re(\vec{u}) + i\ Im(\vec{u})\big] \\ ARe(\vec{u}) + iAIm(\vec{u}) &= \big[aRe(\vec{u}) - bIm(\vec{u})\big] + i\big[bRe(\vec{u}) + aIm(\vec{u})\big]. \end{aligned}$$

Al igualar las partes real e imaginaria, se llega al sistema de ecuaciones (2.21).

Además, por la Definición 2.1 se tiene que $\vec{u}$ es no nulo, luego si $Im(\vec{u}) = \vec{0}$, entonces $Re(\vec{u}) \neq \vec{0}$, y de la segunda ecuación de (2.21), se tiene que $b\ Re(\vec{u}) = \vec{0}$, es decir $b = 0$, lo cual contradice la suposición de que $b \neq 0$, por lo tanto $Im(\vec{u}) \neq \vec{0}$.

Veamos ahora que $Re(\vec{u})$ y $Im(\vec{u})$ son vectores linealmente independientes (por contradicción). Supongamos que $Re(\vec{u}) = \alpha\ Im(\vec{u})$, si se reemplaza en (2.21), se tiene que

$$\begin{aligned} A\ \alpha\ Im(\vec{u}) &= a\ \alpha\ Im(\vec{u}) - b\ Im(\vec{u}) \\ A\ Im(\vec{u}) &= b\ \alpha\ Im(\vec{u}) + a\ Im(\vec{u}). \end{aligned}$$

Si se resuelve dicho sistema de ecuaciones, se obtiene que

$$\big(\alpha^2 + 1\big) b\ Im\big(\vec{u}\big) = \vec{0}.$$

Como $b \neq 0$ y $Im\big(\vec{u}\big) \neq \vec{0}$, entonces $\alpha = \pm i$, luego $Re\big(\vec{u}\big) \in \mathbb{C}^2$, lo cual es contradictorio ya que $Re\big(\vec{u}\big)$ y $Im\big(\vec{u}\big) \in \mathbb{R}^2$. ■

El corolario que se enuncia a continuación muestra que una matriz con componentes reales cuyos valores propios son complejos no es diagonalizable.

Corolario 2.22.1 *Sea A una matriz real de tamaño 2×2 con un valor propio complejo $\lambda = a + bi$ $(b \neq 0)$ y vector propio asociado $\vec{u} \in \mathbb{C}^2$, entonces:*

$$A = PRP^{-1}, \tag{2.22}$$

donde

$$P = \begin{bmatrix} Re\big(\vec{u}\big) & Im\big(\vec{u}\big) \end{bmatrix} \qquad y \qquad R = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}. \tag{2.23}$$

Demostración.

Expresando el sistema de ecuaciones propuesto en (2.21) en forma matricial, se tiene que

$$A \begin{bmatrix} Re\big(\vec{u}\big) & Im\big(\vec{u}\big) \end{bmatrix} = \begin{bmatrix} Re\big(\vec{u}\big) & Im\big(\vec{u}\big) \end{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix},$$

es decir, $AP = PR$. Pero en el Teorema 2.22, se demostró que $Re(\vec{u})$ y $Im(\vec{u})$ eran vectores linealmente independientes, luego P es no singular. Por lo tanto,

$$A = PRP^{-1}. \ ■$$

Ejemplo 2.16 *Expresar de la forma PRP^{-1} la matriz:*

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}.$$

Solución.

En el Ejemplo 2.14, se encontraron los dos valores propios $\lambda_1 = 1 + i\sqrt{2}$ y $\lambda_2 = 1 - i\sqrt{2}$, y los respectivos vectores propios $\vec{v}_1 = \begin{bmatrix} 1 \\ -i\sqrt{2} \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 1 \\ i\sqrt{2} \end{bmatrix}$. Estableciendo:

$$P = \begin{bmatrix} 1 & 0 \\ 0 & -\sqrt{2} \end{bmatrix} \qquad \text{y} \qquad R = \begin{bmatrix} 1 & \sqrt{2} \\ -\sqrt{2} & 1 \end{bmatrix},$$

el lector puede verificar fácilmente que $A = PRP^{-1}$.

Ejemplo 2.17 *Encuentre las matrices P y R, de tal manera que se pueda expresar la siguiente matriz A como PRP^{-1} :*

$$A = \begin{bmatrix} 29 & 9 & -31 \\ 20 & 70 & -5 \\ 66 & -164 & 51 \end{bmatrix}.$$

Solución.

La ecuación característica de A está dada por

$$\begin{aligned} \det\big(A - \lambda I\big) &= -\lambda^3 + 150\lambda^2 - 8125\lambda + 312500 \\ &= -(\lambda - 100)(\lambda - 25 + 50\,i)(\lambda - 25 - 50\,i) = 0. \end{aligned}$$

Por lo tanto, los valores propios son $\lambda_1 = 100$, $\lambda_2 = 25 - 50\,i$ y $\lambda_3 = \overline{\lambda_2}$.

Para $\lambda_1 = 100$, se resuelve $(A - \lambda_1 I)\vec{v} = \vec{0}$ y se obtiene el vector propio asociado $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$. Para $\lambda_2 = 25 - 50\,i$, se resuelve $(A - \lambda_2 I)\vec{v} = \vec{0}$ y se obtiene el vector propio complejo asociado $\vec{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \\ -2 \end{bmatrix} i$.

Por consiguiente, estableciendo

$$P = \begin{bmatrix} -1 & -3 & 1 \\ 1 & 0 & 1 \\ 5 & -2 & -2 \end{bmatrix} \qquad \text{y} \qquad R = \begin{bmatrix} 25 & -50 & 0 \\ 50 & 25 & 0 \\ 0 & 0 & 100 \end{bmatrix}$$

se obtiene

$$\begin{aligned} PRP^{-1} &= -\frac{1}{25}\begin{bmatrix} -1 & -3 & 1 \\ 1 & 0 & 1 \\ 5 & -2 & -2 \end{bmatrix}\begin{bmatrix} 25 & -50 & 0 \\ 50 & 25 & 0 \\ 0 & 0 & 100 \end{bmatrix}\begin{bmatrix} 2 & -8 & -3 \\ 7 & -3 & 2 \\ -2 & -17 & 3 \end{bmatrix} \\ &= -\frac{1}{25}\begin{bmatrix} -1 & -3 & 1 \\ 1 & 0 & 1 \\ 5 & -2 & -2 \end{bmatrix}\begin{bmatrix} -300 & -50 & -175 \\ 275 & -475 & -100 \\ -200 & -1700 & 300 \end{bmatrix} \\ &= -\frac{1}{25}\begin{bmatrix} -725 & -225 & 775 \\ -500 & -1750 & 125 \\ -1650 & 4100 & -1275 \end{bmatrix} = \begin{bmatrix} 29 & 9 & -31 \\ 20 & 70 & -5 \\ 66 & -164 & 51 \end{bmatrix}. \end{aligned}$$

En este ejemplo se ilustra la manera de expresar A como PRP^{-1} cuando sus valores propios no son todos complejos.

Ejercicios 2.3

1. Exprese cada una de las matrices dadas como PRP^{-1}

$$a.\ \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}. \qquad b.\ \begin{bmatrix} 1 & -1 \\ 3 & -1 \end{bmatrix}. \qquad c.\ \begin{bmatrix} -1 & 3 \\ -1 & 1 \end{bmatrix}.$$

$$d.\ \begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \qquad e.\ \begin{bmatrix} 2 & 2 & -3 \\ -3 & 0 & 3 \\ 4 & 1 & -5 \end{bmatrix}. \qquad f.\ \begin{bmatrix} -5 & 2 & 4 \\ -3 & 0 & 3 \\ -3 & 1 & 2 \end{bmatrix}.$$

2. Suponga que A es una matriz real de tamaño 3×3 tal que $\det A = 50$, $\mathrm{tr}_1(A) = 8$ y un valor propio es 2. Encuentre los valores propios.

3. Sea A una matriz real de tamaño $n \times n$ y sea $\vec{x} \in \mathbb{C}^n$. Muestre que

$$Re(A\vec{x}) = ARe(\vec{x}) \qquad \text{y} \qquad Im(A\vec{x}) = AIm(\vec{x}).$$

2.4 Diagonalización de matrices simétricas

Como se vio en la sección anterior, una matriz A de tamaño $n \times n$ puede tener valores propios complejos, aun en el caso de que todos los elementos de A sean reales. En este apartado se desarrollará la teoría de valores propios para matrices simétricas reales.

Definición 2.11 Matrices congruentes ortogonalmente

Dos matrices simétricas A y B de tamaño $n \times n$ son congruentes ortogonalmente si existe una matriz P ortogonal de tamaño $n \times n$ tal que

$$A = P^t BP. \tag{2.24}$$

Definición 2.12 Matrices semejantes ortogonalmente

Una matriz simétrica A de tamaño $n \times n$ es semejante ortogonalmente a una matriz simétrica B de tamaño $n \times n$ si existe una matriz P ortogonal de tamaño $n \times n$ tal que

$$A = P^t BP. \tag{2.25}$$

Teorema 2.23 *Dos matrices simétricas reales A y B son congruentes ortogonalmente si y solo si A y B son semejantes ortogonalmente.*

Demostración.
Si A y B son matrices simétricas congruentes de tamaño $n \times n$, entonces

$$B = P^t AP,$$

pero como $P^t P = I_n$, se tiene que $P^t = P^{-1}$. Por lo tanto, las matrices A y B son semejantes ortogonalmente. ■

Teorema 2.24 *Sea A una matriz simétrica real de tamaño $n \times n$, entonces los valores propios de A son reales.*

Demostración.

Sea $\vec{v} \in \mathbb{C}^n$ un vector propio asociado al valor propio λ de A, entonces

$$A\vec{v} = \lambda\vec{v}. \tag{2.26}$$

Por el Teorema 2.21, se tiene que $\overline{\vec{v}}$ es también vector propio de A, pero asociado al valor propio $\overline{\lambda}$.

Si se multiplica (2.26) por la izquierda por $\overline{\vec{v}}^{\,t}$ se obtiene

$$\overline{\vec{v}}^{\,t} A\vec{v} = \overline{\vec{v}}^{\,t} \lambda\vec{v} = \lambda \overline{\vec{v}}^{\,t} \vec{v}. \tag{2.27}$$

Luego, la conjugada de $\overline{\vec{v}}^{\,t} A\vec{v}$ es

$$\overline{\overline{\vec{v}}^{\,t} A\vec{v}} = \overline{\overline{\vec{v}}^{\,t}}\ \overline{A}\ \overline{\vec{v}} = \overline{\overline{\vec{v}}^{\,t}}\ \overline{\lambda}\ \overline{\vec{v}} = \overline{\lambda} \vec{v}^{\,t} \overline{\vec{v}}, \tag{2.28}$$

donde hemos utilizado el Teorema 1.42. Por otra parte, como A es real, se tiene que $A = \overline{A}$. Por lo tanto, la ecuación (2.28) es igual a

$$\overline{\overline{\vec{v}}^{\,t} A\vec{v}} = \overline{\overline{\vec{v}}^{\,t}}\ A\ \overline{\vec{v}} = \vec{v}^{\,t} A^t \overline{\vec{v}} = (A\vec{v})^{\,t} \overline{\vec{v}} = \lambda \vec{v}^{\,t} \overline{\vec{v}}, \tag{2.29}$$

aquí se utilizó el hecho de que $A^t = A$, ya que A es simétrica.

Si se igualan (2.28) y (2.29), se tiene

$$\overline{\lambda}\ \vec{v}^{\,t}\ \overline{\vec{v}} = \lambda\ \vec{v}^{\,t}\ \overline{\vec{v}}, \tag{2.30}$$

pero $\vec{v}^{\,t}\ \overline{\vec{v}} = \|\vec{v}\|^2 \neq 0$, ya que $\vec{v}$ es un vector propio. Entonces se puede dividir ambos lados de (2.30) entre $\vec{v}^t\ \overline{\vec{v}}$ para obtener

$$\overline{\lambda} = \lambda,$$

lo cual se cumple solo si λ es real. ■

Definición 2.13 Matriz diagonalizable ortogonalmente

Una matriz A de tamaño $n \times n$ es diagonalizable ortogonalmente si existe una matriz ortogonal Q tal que

$$Q^t A Q = D \tag{2.31}$$

sea diagonal.

Teorema 2.25 *Si A es una matriz simétrica real de tamaño $n \times n$, entonces existe una matriz ortogonal Q tal que*

$$Q^{-1}AQ$$

es una matriz diagonal.

Demostración.

Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ los valores propios de A. Puesto que λ_1 es real, existe un vector propio unitario $\vec{u}_1 \in \mathbb{R}^n$ de A correspondiente a λ_1. Denotemos por $\mathbb{V}$ el complemento ortogonal a $\vec{u}_1$ de dimensión $n-1$. Sea $\{\vec{u}_2, \vec{u}_3, \ldots, \vec{u}_n\}$ una base ortonormal de $\mathbb{V}$. Luego, cada vector $\vec{X}$ de $\mathbb{V}$ tiene la forma

$$\vec{X} = a_2\ \vec{u}_2 + a_3\ \vec{u}_3 + \ldots + a_n\ \vec{u}_n,$$

y el producto punto entre $A\vec{X}$ y $\vec{u}_1$ es

$$(A\vec{X}) \cdot \vec{u}_1 = (A\vec{X})^t\vec{u}_1 = \vec{X}^t A^t\vec{u}_1 = \vec{X}^t(A\vec{u}_1) = \vec{X}^t(\lambda_1\vec{u}_1) = \lambda_1\vec{X}^t\vec{u}_1 = 0,$$

puesto que cada vector de la base de $\mathbb{V}$ es ortogonal a $\vec{u}_1$. La matriz de cambio de base de la *base canónica* de $\mathbb{R}^n$ a la base $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ es la matriz ortogonal S cuyas columnas son los elementos de los vectores $\vec{u}_i$. Luego

$$\begin{aligned} AS &= \begin{bmatrix} A\vec{u}_1 & A\vec{u}_2 & \ldots & A\vec{u}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\vec{u}_1 & A\vec{u}_2 & \ldots & A\vec{u}_n \end{bmatrix}, \end{aligned}$$

por lo tanto

$$S^{-1}AS = S^{-1}\begin{bmatrix} \lambda_1\vec{u}_1 & A\vec{u}_2 & \ldots & A\vec{u}_n \end{bmatrix}.$$

Pero como S es ortogonal, se tiene que $S^{-1} = S^t$, por consiguiente

$$\begin{aligned} S^{-1}AS &= \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{bmatrix}^t \begin{bmatrix} \lambda_1\vec{u}_1 & A\vec{u}_2 & \ldots & A\vec{u}_n \end{bmatrix} \\ &= \begin{bmatrix} \vec{u}_1^t\lambda_1\vec{u}_1 & \vec{u}_1^tA\vec{u}_2 & \ldots & \vec{u}_1^tA\vec{u}_n \\ \vec{u}_2^t\lambda_1\vec{u}_1 & \vec{u}_2^tA\vec{u}_2 & \ldots & \vec{u}_2^tA\vec{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{u}_n^t\lambda_1\vec{u}_1 & \vec{u}_n^tA\vec{u}_2 & \ldots & \vec{u}_n^tA\vec{u}_n \end{bmatrix}. \end{aligned}$$

Se puede probar fácilmente que $S^{-1}AS$ es simétrica, ya que

$$(S^{-1}AS)^t = (S^tAS)^t = S^tAS = S^{-1}AS.$$

Por consiguiente

$$S^{-1}AS = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{bmatrix},$$

donde A_1 es simétrica de tamaño $(n-1)\times(n-1)$.

La prueba se completa por inducción, si R^* es una matriz ortogonal de tamaño $(n-1)\times(n-1)$ tal que $R^*A_1(R^*)^{-1} = diag\{\lambda_2, \lambda_3, \dots, \lambda_n\}$. Entonces, la matriz

$$R = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & R^* & \\ 0 & & & \end{bmatrix}$$

es una matriz ortogonal y

$$R^{-1}S^{-1}ASR = (SR)^{-1}A(SR) = diag\{\lambda_1, \lambda_2, \dots, \lambda_n\}.$$

La matriz $SR = Q$ es el producto de dos matrices ortogonales, por lo tanto es también una matriz ortogonal.

Así, $Q^{-1}AQ$ es una matriz diagonal y la prueba queda completa. ■

Teorema 2.26 *Sea A una matriz simétrica real de tamaño $n \times n$. Entonces los vectores propios asociados con valores propios distintos de A son ortogonales. Esto es, los espacios propios de una matriz simétrica son ortogonales.*

Demostración.

Sean $\vec{v}_1$ y $\vec{v}_2$ vectores propios que corresponden a valores propios distintos, digamos λ_1 y λ_2. Para demostrar que $\vec{v}_1 \cdot \vec{v}_2 = 0$, se calcula

$$\begin{aligned} \lambda_1\vec{v}_1 \cdot \vec{v}_2 &= (\lambda_1\vec{v}_1)^t\vec{v}_2 = (A\vec{v}_1)^t\vec{v}_2 && \text{puesto que } \vec{v}_1 \text{ es un vector propio} \\ &= (\vec{v}_1^t\ A^t)\vec{v}_2 = \vec{v}_1^t(A\vec{v}_2) && \text{puesto que } A \text{ es simétrica} \\ &= \vec{v}_1^t(\lambda_2\vec{v}_2) && \text{puesto que } \vec{v}_2 \text{ es un vector propio} \\ &= \lambda_2\vec{v}_1^t\vec{v}_2 = \lambda_2\vec{v}_1 \cdot \vec{v}_2. \end{aligned}$$

Por lo tanto, $(\lambda_1 - \lambda_2)\vec{v}_1 \cdot \vec{v}_2 = 0$. Pero $\lambda_1 - \lambda_2 \neq 0$, así que $\vec{v}_1 \cdot \vec{v}_2 = 0$. ■

Teorema 2.27 *Sea λ_k un valor propio de multiplicidad algebraica igual a p asociado a una matriz A simétrica real de tamaño $n \times n$. Entonces A tiene exactamente p vectores propios mutuamente ortogonales asociados al valor propio λ_k.*

Demostración.

Por el Teorema 2.25, existe una matriz Q tal que $Q^{-1}AQ$ es una matriz diagonal en la cual λ_k aparece exactamente p veces en la diagonal principal. Por otra parte, se tiene que $Q^{-1}AQ - \lambda_k I_n = Q^{-1}\big(A - \lambda_k I_n\big)Q$ tiene rango $n - p$. Pero como Q y Q^{-1} son no singulares, $A - \lambda_k I_n$ también tiene rango $n - p$. Por lo tanto, el espacio solución del sistema de ecuaciones

$$\big(A - \lambda_k I_n\big)\vec{v} \;=\; \vec{0}, \qquad\qquad \text{con} \qquad\qquad \vec{v} \in \mathbb{R}^n$$

tiene dimensión $n - (n - p) = p$ y, por consiguiente, existen exactamente p vectores unitarios mutuamente ortogonales de $\mathbb{R}^n$. ■

Teorema 2.28 *Una matriz A simétrica real de tamaño $n \times n$ tiene n vectores propios unitarios mutuamente ortogonales.*

Demostración.

Si D es la matriz diagonal Q^tAQ, se tiene

$$AQ = QD, \tag{2.32}$$

donde $D = diag\{\lambda_1, \ldots, \lambda_n\}$. Al igualar los vectores columna de cada miembro de (2.32), se obtiene

$$A\ \vec{v}_1 = \lambda_1\ \vec{v}_1,\ A\ \vec{v}_2 = \lambda_2\ \vec{v}_2,\ \ldots, A\ \vec{v}_n = \lambda_n\ \vec{v}_n,$$

donde $\vec{v}_1,\ \vec{v}_2,\ \ldots,\ \vec{v}_n$ son los vectores columna de Q. Se deduce que las columnas de Q son vectores propios de A y que son vectores unitarios mutuamente ortogonales, por ser Q ortogonal. ■

Teorema 2.29 *Sea A una matriz real de tamaño $n \times n$. Entonces A es diagonalizable ortogonalmente si y solo si A es simétrica.*

Demostración.

Sea A diagonalizable ortogonalmente, entonces por la Definición 2.13, existe una matriz ortogonal Q tal que $Q^tAQ = D$. Si se multiplica esta ecuación por la izquierda por Q y por la derecha por Q^t y se usa el hecho de que $Q^tQ = QQ^t = I_n$, se obtiene

$$A = QDQ^t.$$

Luego

$$A^t = \left(QDQ^t\right)^t = \left(Q^t\right)^t D^t Q^t = QDQ^t = A.$$

Así, A es simétrica.

Recíprocamente, suponga que A es simétrica. Entonces por los Teoremas 2.26 y 2.28, A es diagonalizable ortogonalmente con la matriz Q cuyas columnas son los vectores propios dados en el Teorema 2.28, y el teorema queda demostrado. ■

Procedimiento para encontrar una matriz diagonalizante Q

i) Encuentre una base para cada espacio propio de A.

ii) Halle una base ortonormal para cada espacio propio de A usando el proceso de Gram-Schmidt o algún otro.

iii) Obtenga Q como la matriz cuyas columnas son los vectores propios ortonormales obtenidos en el paso ii).

Ejemplo 2.18 *Encuentre una matriz Q que diagonalice ortogonalmente a la siguiente matriz:*

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

Solución.

En este caso, los valores propios de A son $\lambda_1 = 1$ (de multiplicidad algebraica 2) y $\lambda_2 = 7$. Los vectores propios linealmente independientes

correspondientes a $\lambda_1 = 1$ son $\vec{v}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$, y el correspondiente a $\lambda_2 = 7$ es $\vec{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

Para encontrar Q, se aplica el proceso de Gram-Schmidt a $\{\vec{v}_1, \vec{v}_2\}$ una base para $\mathcal{B}_1$. Como $\|\vec{v}_1\| = \sqrt{2}$, se hace $\vec{u}_1 = \dfrac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$. Después,

$$\begin{aligned} \vec{v}_2' &= \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} - \frac{-2}{\sqrt{2}} \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Entonces, $\|\vec{v}_2'\| = \sqrt{3}$ y $\vec{u}_2 = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$. Se puede verificar que la nueva base de $\mathcal{B}_1$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_2 = 0$. Por último, se tiene que $\|\vec{v}_3\| = \sqrt{6}$, luego $\vec{u}_3 = \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$. También se verifica que la base obtenida para $\mathbb{R}^3$ es ortonormal, observando que $\vec{u}_1 \cdot \vec{u}_3 = 0$ y $\vec{u}_2 \cdot \vec{u}_3 = 0$. Por lo tanto,

$$Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}, \quad Q^t = \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix},$$

y

$$\begin{aligned} Q^tAQ &= \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \\ &= \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 7/\sqrt{6} \\ 0 & -1/\sqrt{3} & 14/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 7/\sqrt{6} \end{bmatrix}. \end{aligned}$$

Así,

$$Q^tAQ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}.$$

El siguiente teorema es un caso particular del Teorema 2.13.

Teorema 2.30 Descomposición espectral de matrices simétricas

Sea A una matriz simétrica real de tamaño $n \times n$ con valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces A se puede escribir como

$$A = \sum_{i=1}^{n} \lambda_i \mathbf{E}(\lambda_i), \tag{2.33}$$

donde las matrices de proyección espectral vienen dadas por

$$\mathbf{E}(\lambda_i) = \frac{\vec{u}_i \vec{u}_i^t}{\langle \vec{u}_i, \vec{u}_i \rangle} = \frac{\vec{u}_i \vec{u}_i^t}{\vec{u}_i^t \vec{u}_i} = \vec{v}_i \vec{v}_i^t, \tag{2.34}$$

con $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ los vectores propios normalizados de A.

Demostración.

Por el Teorema 2.25, existe una matriz Q tal que $Q^{-1}AQ$ es una matriz diagonal. Entonces,

$$\begin{aligned} A &= QDQ^{-1} = QDQ^t \\ &= \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{v}_1^t \\ \vec{v}_2^t \\ \vdots \\ \vec{v}_n^t \end{bmatrix} \\ &= \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 \vec{v}_1^t \\ \lambda_2 \vec{v}_2^t \\ \vdots \\ \lambda_n \vec{v}_n^t \end{bmatrix} \\ &= \lambda_1 \vec{v}_1 \vec{v}_1^t + \lambda_2 \vec{v}_2 \vec{v}_2^t + \ldots + \lambda_n \vec{v}_n \vec{v}_n^t, \end{aligned}$$

lo cual prueba el teorema. ■

Ejemplo 2.19 *Ilustrar el teorema de descomposición espectral para la matriz dada en el Ejemplo 2.18.*

Solución.

Del Ejemplo 2.18, se tiene que los valores propios asociados a la matriz A son $\lambda_1 = 7,\ \lambda_2 = \lambda_3 = 1$ y los respectivos vectores propios normalizados de A eran

$$\vec{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}.$$

Luego, las matrices de proyección espectral $\mathbf{E}(\lambda_k)$ son

$$\mathbf{E}(\lambda_1) = \vec{v}_1\vec{v}_1^t = \frac{1}{6}\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}1 & 2 & 1\end{bmatrix} = \frac{1}{6}\begin{bmatrix}1 & 2 & 1\\2 & 4 & 2\\1 & 2 & 1\end{bmatrix}$$

$$\mathbf{E}(\lambda_2) = \vec{v}_2\vec{v}_2^t = \frac{1}{3}\begin{bmatrix}1\\-1\\1\end{bmatrix}\begin{bmatrix}1 & -1 & 1\end{bmatrix} = \frac{1}{3}\begin{bmatrix}1 & -1 & 1\\-1 & 1 & -1\\1 & -1 & 1\end{bmatrix}$$

$$\mathbf{E}(\lambda_3) = \vec{v}_3\vec{v}_3^t = \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}\begin{bmatrix}1 & 0 & -1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1 & 0 & -1\\0 & 0 & 0\\-1 & 0 & 1\end{bmatrix}.$$

Nótese que $\mathbf{E}(\lambda_1) + \mathbf{E}(\lambda_2) + \mathbf{E}(\lambda_3) = I_3$ y

$$\sum_{i=1}^{3}\lambda_i\mathbf{E}(\lambda_i) = \frac{1}{6}\begin{bmatrix}12 & 12 & 6\\12 & 30 & 12\\6 & 12 & 12\end{bmatrix},$$

la cual coincide con la matriz A dada en el Ejemplo 2.18.

Teorema 2.31 Teorema espectral para matrices simétricas

Sea A una matriz simétrica real de tamaño $n \times n$, entonces las siguientes afirmaciones son equivalentes:

i) A tiene n valores propios reales, contando multiplicidades.

ii) Si λ es un valor propio de A con multiplicidad algebraica k, entonces el espacio propio para λ es k-dimensional.

iii) Los vectores propios correspondientes a valores propios distintos son ortogonales, es decir los espacios propios son mutuamente ortogonales.

iv) A es diagonalizable ortogonalmente.

Demostración.
Queda como ejercicio para el lector. ■

Ejercicios 2.4

1. Determine si las matrices dadas a continuación son diagonalizables ortogonalmente:

$$a.\begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}. \qquad b.\begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}. \qquad c.\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

$$d.\begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{bmatrix}. \qquad e.\begin{bmatrix} 3 & -1 & 4 \\ -1 & 1 & -5 \\ 4 & -5 & 2 \end{bmatrix}.$$

2. Si las matrices A y B son ortogonalmente semejantes y B es ortogonalmente semejante a una matriz C, muestre que A y C son también ortogonalmente semejantes.

3. Sea A una matriz ortogonal de tamaño $n \times n$, muestre que si λ es un valor propio de A, entonces $|\lambda| = 1$.

4. Muestre que si A es ortogonal de tamaño $n \times n$ y $\vec{x}$ y $\vec{y}$ son vectores en $\mathbb{R}^n$, entonces $(A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y}$.

2.5 Vectores propios generalizados

En las secciones anteriores hemos considerado matrices en las cuales la multiplicidad algebraica de cada valor propio es igual a su multiplicidad geométrica. En este apartado consideraremos matrices que violan esta condición, es decir que la multiplicidad algebraica de cada valor propio es diferente de su multiplicidad geométrica y se obtendrá un nuevo concepto de vector propio asociado a la matriz.

Definición 2.14 Vector propio generalizado

Sea A una matriz real de tamaño $n \times n$ con un valor propio λ_j cuya multiplicidad algebraica es diferente de su multiplicidad geométrica. Un vector $\vec{v} \in \mathbb{C}^n$ se llama vector propio generalizado de A si cumple que

$$\left(A - \lambda_j I_n\right)^k \vec{v} = \vec{0} \tag{2.35}$$

para algún k entero positivo. El mínimo entero k para el cual (2.35) se satisface recibe el nombre de índice del vector propio generalizado $\vec{v}$.

Nota 2.3 Los vectores propios de A son vectores propios generalizados de índice igual a 1.

Ejemplo 2.20 *Verifique si el vector $\vec{v}^t = (-\frac{1}{7}, 0)$ es un vector propio generalizado asociado al valor propio $\lambda = -5$ de la siguiente matriz:*

$$A = \begin{bmatrix} -12 & 7 \\ -7 & 2 \end{bmatrix}.$$

Solución.

Veamos que $\vec{v}$ cumple (2.35) para algún valor entero k. Para $k = 1$,

$$\big[A - (\text{-}5)I\big]\vec{v} = \begin{bmatrix} \text{-}7 & 7 \\ \text{-}7 & 7 \end{bmatrix} \begin{bmatrix} -\frac{1}{7} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \vec{v}_1.$$

Obsérvese que $\vec{v}_1$ es el vector propio correspondiente a λ. Para $k = 2$, se tiene que

$$\big[A - (\text{-}5)I\big]^2\vec{v} = \big[A - (\text{-}5)I\big]\vec{v}_1 = \begin{bmatrix} \text{-}7 & 7 \\ \text{-}7 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Luego, $\vec{v}$ es un vector propio generalizado de índice $k = 2$.

Definición 2.15 Espacio propio generalizado

Sea λ_j un valor propio de la matriz real $A \in \mathcal{M}_{nn}$. El subespacio

$$\mathcal{V}_j = \{\vec{v} \in \mathbb{C}^n : \big(A - \lambda_j I\big)^k \vec{v} = \vec{0}, \textit{para cierto entero positivo } k\} \quad (2.36)$$

se denomina espacio propio generalizado de A asociado con el valor propio λ_j. En otras palabras, $\mathcal{V}_j = \ker\big\{\big(A - \lambda_j I\big)^k\big\}$.

Teorema 2.32 *Sea A una matriz real de tamaño 2×2 con un único valor propio real λ de multiplicidad algebraica distinta de la multiplicidad geométrica. Entonces existe un vector propio generalizado $\vec{w}$ que satisface la ecuación:*

$$(A - \lambda I)\vec{w} = \vec{v}, \quad (2.37)$$

donde $\vec{v}$ es un vector propio correspondiente a λ.

Demostración.

Sea $\vec{x} \in \mathbb{C}^2$ un vector fijo tal que $\vec{x} \neq \alpha\vec{v}$. Luego, $\vec{x}$ no es un vector propio de A. Sea

$$\vec{y} = (A - \lambda I)\vec{x}. \quad (2.38)$$

Demostremos que $\vec{y}$ es un vector propio de A; en otras palabras, que $\vec{y} = \beta\vec{v}$. Como $\vec{v}$ y $\vec{x}$ son linealmente independientes y $\vec{y} \in \mathbb{C}^2$, entonces existen constantes c_1 y c_2 tales que

$$\vec{y} = c_1\ \vec{v} + c_2\ \vec{x}. \tag{2.39}$$

Se debe mostrar que $c_2 = 0$. Si se reemplaza (2.38) en (2.39), se tiene que

$$\begin{aligned}(A - \lambda I)\vec{x} &= c_1\ \vec{v} + c_2\ \vec{x}\\ \big[A - (\lambda + c_2)I\big]\vec{x} &= c_1\ \vec{v}.\end{aligned}$$

Si se supone que $c_2 \neq 0$, entonces $\lambda + c_2$ no es un valor propio de A (el único valor propio asociado a A es λ). Por lo tanto,

$$\det\big[A - (\lambda + c_2)I\big] \neq 0.$$

Sea $B = A - (\lambda + c_2)I$, entonces B es no singular. Así, $\vec{x}$ es igual a

$$\vec{x} = B^{-1}\ c_1\ \vec{v} = c_1\ B^{-1}\ \vec{v}. \tag{2.40}$$

Al multiplicar a ambos lados de (2.40) por λ, se obtiene

$$\lambda\ \vec{x} = \lambda\ B^{-1}\ c_1\ \vec{v} = c_1\ B^{-1}\ (\lambda\ \vec{v}) = c_1\ B^{-1}\ (A\ \vec{v}).$$

Pero $A = B + (\lambda + c_2)I$, de manera que

$$\begin{aligned}\lambda\ \vec{x}\ &= c_1\ B^{-1}\big[B + (\lambda + c_2)I\big]\ \vec{v}\\ &= c_1\ \big[I + (\lambda + c_2)B^{-1}\big]\ \vec{v}\\ &= c_1\ \vec{v} + (\lambda + c_2)\big[c_1\ B^{-1}\ \vec{v}\big]\\ &= c_1\ \vec{v} + (\lambda + c_2)\ \vec{x},\end{aligned}$$

la cual se obtiene usando el hecho de que $\vec{x} = c_1\ B^{-1}\ \vec{v}$. Por lo tanto,

$$\begin{aligned}\lambda\ \vec{x}\ &= c_1\ \vec{v} + \lambda\ \vec{x} + c_2\ \vec{x}\\ \vec{0}\ &= c_1\ \vec{v} + c_2\ \vec{x}.\end{aligned}$$

Pero como $\vec{x} \neq \alpha\vec{v}$, se debe tener que $c_1 = c_2 = 0$, lo cual contradice la suposición de que $c_2 \neq 0$. Luego, $c_2 = 0$ y sustituyendo en (2.39), se tiene que $\vec{y} = c_1\vec{v}$.

Ahora, se debe comprobar que $c_1 \neq 0$, en otras palabras se debe mostrar que $\vec{y} \neq \vec{0}$; pues si $\vec{y} = \vec{0}$, al reemplazar en (2.38) se tendría que $\vec{x}$ es un vector propio de A, lo cual contradice la suposición de que

$\vec{x} \neq \alpha\vec{v}$. Luego, $c_1 \neq 0$, es decir $\vec{y}$ es un múltiplo no nulo de $\vec{v}$ y por el Teorema 2.1, es un vector propio de A.

Por último definamos $\vec{w} = \frac{1}{c_1}\vec{x}$, entonces

$$(A - \lambda I)\vec{w} = \frac{1}{c_1}(A - \lambda I)\vec{x} = \frac{1}{c_1}\vec{y} = \vec{v}.$$

Esto prueba el teorema. ■

Ejemplo 2.21 *Sea* $A = \begin{bmatrix} 4 & -1 \\ 1 & 2 \end{bmatrix}$. *Determine sus vectores propios generalizados.*

Solución.

La ecuación característica de A es

$$\lambda^2 - 6\lambda + 9 = (\lambda - 3)^2 = 0.$$

Luego, $\lambda = 3$ es el único valor propio (de multiplicidad algebraica 2), entonces

$$\big(A - \lambda I\big)\vec{v} = (A - 3I)\vec{v} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Esto conduce a que $x_1 = x_2$. Estableciendo $x_2 = 1$, se obtiene solo un vector propio linealmente independiente: $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Luego, para encontrar un vector propio generalizado $\vec{v}_2$, se calcula $(A - 3I)\vec{v}_2 = \vec{v}_1$ y se obtiene

$$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

La solución general corresponde a $x_1 - x_2 = 1$, luego $x_1 = 1 + x_2$. Por lo tanto si $x_2 = 0$, se tiene $\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Ejemplo 2.22 *Encuentre los vectores propios generalizados de la matriz*

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ -3 & 2 & 4 \end{bmatrix}.$$

Solución.

La ecuación característica de A es

$$-\lambda^3 + 6\lambda^2 - 12\lambda + 8 = -\big(\lambda - 2\big)^3 = 0.$$

Luego, $\lambda = 2$ es el único valor propio (de multiplicidad algebraica 3), entonces

$$(A - \lambda I)\vec{v} = (A - 2I)\vec{v} = \begin{bmatrix} \text{-}1 & 1 & 1 \\ 2 & \text{-}1 & \text{-}1 \\ \text{-}3 & 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Por lo tanto, el vector propio correspondiente se obtiene operando por filas

$$\left[\begin{array}{ccc|c} \text{-}1 & 1 & 1 & 0 \\ 2 & \text{-}1 & \text{-}1 & 0 \\ \text{-}3 & 2 & 2 & 0 \end{array}\right] \underset{\substack{F_2+2F_1 \\ F_3+F_2}}{\sim} \left[\begin{array}{ccc|c} \text{-}1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ \text{-}1 & 1 & 1 & 0 \end{array}\right] \underset{\substack{F_3-F_1 \\ F_1-F_2}}{\sim} \left[\begin{array}{ccc|c} \text{-}1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Esto conduce a que $x = 0$ y $y = -z$. Estableciendo $z = -1$, se obtiene el vector propio: $\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ \text{-}1 \end{bmatrix}$. Luego, para encontrar un vector propio generalizado $\vec{v}_2$, se calcula $(A - 2I)\vec{v}_2 = \vec{v}_1$ y se tiene que

$$\begin{bmatrix} \text{-}1 & 1 & 1 \\ 2 & \text{-}1 & \text{-}1 \\ \text{-}3 & 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ \text{-}1 \end{bmatrix},$$

y si se realizan operaciones por filas se obtiene

$$\left[\begin{array}{ccc|c} \text{-}1 & 1 & 1 & 0 \\ 2 & \text{-}1 & \text{-}1 & 1 \\ \text{-}3 & 2 & 2 & \text{-}1 \end{array}\right] \underset{\substack{F_2+2F_1 \\ F_3+F_2}}{\sim} \left[\begin{array}{ccc|c} \text{-}1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ \text{-}1 & 1 & 1 & 0 \end{array}\right] \underset{\substack{F_3-F_1 \\ F_1-F_2}}{\sim} \left[\begin{array}{ccc|c} \text{-}1 & 0 & 0 & \text{-}1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Es decir, $x = 1$ y $y = 1 - z$. Si se hace $z = 0$, se obtiene el vector propio generalizado: $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Para encontrar el segundo vector propio generalizado $\vec{v}_3$, se calcula

$$(A - 2I)\vec{v}_3 = \vec{v}_2$$

$$\begin{bmatrix} \text{-}1 & 1 & 1 \\ 2 & \text{-}1 & \text{-}1 \\ \text{-}3 & 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Al resolver el sistema por Gauss-Jordan, se tiene que

$$\left[\begin{array}{ccc|c} -1 & 1 & 1 & 1 \\ 2 & -1 & -1 & 1 \\ -3 & 2 & 2 & 0 \end{array}\right] \underset{\substack{F_2+2F_1 \\ F_3+F_2}}{\sim} \left[\begin{array}{ccc|c} -1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 3 \\ -1 & 1 & 1 & 1 \end{array}\right] \underset{\substack{F_3-F_1 \\ F_1-F_2}}{\sim} \left[\begin{array}{ccc|c} -1 & 0 & 0 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Luego, $x = 2$ y $y = 3 - z$. Si $z = 0$, se obtiene el vector propio generalizado:

$$\vec{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}.$$

Ejemplo 2.23 *Encuentre los vectores propios generalizados de la matriz:*

$$B = \begin{bmatrix} 1 & 18 & 7 \\ -1 & 13 & 4 \\ 1 & -25 & -8 \end{bmatrix}.$$

Solución.

La ecuación característica de B es

$$-\lambda^3 + 6\lambda^2 - 12\lambda + 8 = 0.$$

Luego, $\lambda = 2$ es el único valor propio (de multiplicidad algebraica tres), entonces

$$(B - \lambda I)\vec{v} = (B - 2I)\vec{v} = \begin{bmatrix} -1 & 18 & 7 \\ -1 & 11 & 4 \\ 1 & -25 & -10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

El vector propio correspondiente se obtiene operando por filas

$$\left[\begin{array}{ccc|c} -1 & 18 & 7 & 0 \\ -1 & 11 & 4 & 0 \\ 1 & -25 & -10 & 0 \end{array}\right] \underset{\substack{F_2-F_1 \\ F_3+F_1}}{\sim} \left[\begin{array}{ccc|c} -1 & 18 & 7 & 0 \\ 0 & -7 & -3 & 0 \\ 0 & -7 & -3 & 0 \end{array}\right] \underset{\substack{F_3-F_2 \\ 3F_1+7F_2}}{\sim} \left[\begin{array}{ccc|c} -3 & 5 & 0 & 0 \\ 0 & 7 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Esto conduce a que $x = \frac{5}{3}y$ y $z = -\frac{7}{3}y$. Estableciendo $y = 3$, se obtiene el vector propio: $\vec{v}_1 = \begin{bmatrix} 5 \\ 3 \\ -7 \end{bmatrix}$. Para encontrar un vector propio generalizado

$\vec{v}_2$, se calcula $(B-2I)\vec{v}_2 = \vec{v}_1$ y se tiene que

$$\begin{bmatrix} -1 & 18 & 7 \\ -1 & 11 & 4 \\ 1 & -25 & -10 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ -7 \end{bmatrix}.$$

Si se realizan operaciones por filas, se obtiene

$$\left[\begin{array}{ccc|c} -1 & 18 & 7 & 5 \\ -1 & 11 & 4 & 3 \\ 1 & -25 & -10 & -7 \end{array}\right] \underset{\substack{F_2-F_1 \\ F_3+F_1}}{\sim} \left[\begin{array}{ccc|c} -1 & 18 & 7 & 5 \\ 0 & -7 & -3 & -2 \\ 0 & -7 & -3 & -2 \end{array}\right] \underset{\substack{F_3-F_2 \\ 3F_1+7F_2}}{\sim} \left[\begin{array}{ccc|c} -3 & 5 & 0 & 1 \\ 0 & 7 & 3 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Es decir, $x = -\frac{1}{3} + \frac{5}{3}y$ y $z = \frac{2}{3} - \frac{7}{3}y$. Si se hace $y = 0$, se obtiene el vector propio generalizado: $\vec{v}_2 = \frac{1}{3}\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$. Para encontrar el segundo vector propio generalizado $\vec{v}_3$, se calcula $(B-2I)\vec{v}_3 = \vec{v}_2$, es decir

$$\begin{bmatrix} -1 & 18 & 7 \\ -1 & 11 & 4 \\ 1 & -25 & -10 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}.$$

Al resolver el sistema por Gauss-Jordan, se tiene que

$$\left[\begin{array}{ccc|c} -1 & 18 & 7 & -\frac{1}{3} \\ -1 & 11 & 4 & 0 \\ 1 & -25 & -10 & \frac{2}{3} \end{array}\right] \underset{\substack{F_2-F_1 \\ F_3+F_1}}{\sim} \left[\begin{array}{ccc|c} -1 & 18 & 7 & -\frac{1}{3} \\ 0 & -7 & -3 & \frac{1}{3} \\ 0 & -7 & -3 & \frac{1}{3} \end{array}\right] \underset{\substack{F_3-F_2 \\ 3F_1+7F_2}}{\sim} \left[\begin{array}{ccc|c} -3 & 5 & 0 & \frac{4}{3} \\ 0 & 7 & 3 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Luego, $x = -\frac{4}{9} + \frac{5}{3}y$ y $z = -\frac{1}{9} - \frac{7}{3}y$. Tomando $y = 0$, se obtiene el vector propio generalizado: $\vec{v}_3 = -\frac{1}{9}\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$.

Teorema 2.33 *Sea $\vec{v} \neq \vec{0}$ un vector propio generalizado con índice k de una matriz real A de tamaño $n \times n$. Entonces,*

$$\left\{\vec{v}, (A-\lambda_j I_n)\vec{v}, \ldots, (A-\lambda_j I_n)^{k-1}\vec{v}\right\} \tag{2.41}$$

es un conjunto de vectores linealmente independientes.

Demostración.

Supongamos que $\vec{v} \in \mathbb{C}^n$ es un vector propio generalizado de índice k y que los vectores dados en (2.41) son linealmente dependientes, entonces existen constantes $c_m \neq 0$ tales que

$$\sum_{m=0}^{k-1} c_m \, A_j^m \vec{v} \, = \vec{0}, \qquad \text{con} \qquad A_j \, = A - \lambda_j I_n, \tag{2.42}$$

donde $A_j^0 = I_n$. Sea $f_r(t) = c_0 + c_1 t + \ldots + c_r t^r$ un polinomio[1] diferente de cero de grado $r \leq k-1$. Luego, la combinación lineal dada en (2.42) se puede expresar de la siguiente manera:

$$f_r\big(A_j\big)\vec{v} = \vec{0}.$$

Sea $g(t) = t^k$, luego $g\big(A_j\big)\vec{v} = \vec{0}$ por ser $\vec{v}$ vector propio generalizado de índice k. Si existe un polinomio $h(t) = t^d$ con $d < k$ que sea el máximo común divisor de $f_r(t)$ y $g(t)$, usando el algoritmo de Euclides, $h(t)$ se puede escribir como

$$h(t) = h_1(t) f_r(t) + h_2(t) g(t),$$

donde $h_1(t)$ y $h_2(t)$ son polinomios, y así $h\big(A_j\big) = h_1\big(A_j\big) f_r\big(A_j\big) + h_2\big(A_j\big) g\big(A_j\big)$. Por lo tanto,

$$h\big(A_j\big)\vec{v} \, = \vec{0}.$$

Luego, d sería el índice de $\vec{v}$, lo cual contradice la hipótesis de que k es el índice de $\vec{v}$ y se concluye la prueba. ■

Teorema 2.34 *Los vectores no nulos tomados de espacios propios generalizados distintos son linealmente independientes.*

Demostración.

Sea $A_j = A - \lambda_j I_n$ y $\mathcal{V}_j = \ker\Big\{ A_j^{k_j} \Big\}$ para algún entero k_j, $j = 1, 2, \ldots, r$. Sean $\vec{v}_1, \ldots, \vec{v}_r$ con $\vec{v}_j \in \mathcal{V}_j$, y supongamos que

$$\vec{w} = \sum_{j=1}^{r} \vec{v}_j = \vec{0}. \tag{2.43}$$

[1] Si el lector desconoce este concepto, puede ver Lang (2004, cap. 11).

Se debe mostrar que cada $\vec{v}_j = \vec{0}$. Si se multiplican ambos lados de (2.43) por la matriz $C = A_2^{k_2} A_3^{k_3} \dots A_r^{k_r}$ y se utilizan los hechos de que $A_j^{k_j} \vec{v}_j = \vec{0}$ y que $A_i^{k_i} A_j^{k_j} = A_j^{k_j} A_i^{k_i}$, lo cual se tiene ya que

$$\begin{aligned} A_i A_j &= \big(A - \lambda_i I_n\big)\big(A - \lambda_j I_n\big) = A^2 - \lambda_i A - \lambda_j A + \lambda_i \lambda_j I_n \\ &= A^2 - \lambda_j A - \lambda_i A + \lambda_i \lambda_j I_n = \big(A - \lambda_j I_n\big)\big(A - \lambda_i I_n\big) = A_j A_i, \end{aligned}$$

se obtiene que

$$C\vec{w} = C\vec{v}_1 = \vec{0}. \tag{2.44}$$

Por lo tanto, $\vec{v}_1 = \vec{0}$. De manera análoga, todos los $\vec{v}_j$ restantes tienen que desaparecer. ■

De acuerdo con la Definición 2.8, las matrices reales cuadradas A y B se dicen que son *semejantes* si existe una matriz P no singular tal que

$$A = PBP^{-1}.$$

En ocasiones, además de establecer el hecho en sí de la semejanza, se requiere encontrar la matriz P de la transformación que satisface que $A = PBP^{-1}$. En estos momentos, se puede construir la matriz P utilizando los vectores propios generalizados de la siguiente manera

$$P = SR^{-1}, \tag{2.45}$$

donde las columnas de la matriz S son los vectores propios generalizados de la matriz A y las columnas de la matriz R son los vectores propios generalizados de la matriz B.

Ejemplo 2.24 *Determine si las matrices dadas en los Ejemplos 2.22 y 2.23 son semejantes.*

Solución.

Como las dos matrices tienen el mismo polinomio característico, entonces son semejantes. Encontremos la matriz P. En el Ejemplo 2.22, se obtuvieron los siguientes vectores propios generalizados para la matriz A

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}.$$

Del Ejemplo 2.23 se tiene que los vectores propios generalizados de la matriz B son

$$\vec{v}_1 = \begin{bmatrix} 5 \\ 3 \\ -7 \end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{3}\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{9}\begin{bmatrix} -4 \\ 0 \\ -1 \end{bmatrix}.$$

Por lo tanto,

$$P = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 3 \\ -1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 5 & -\frac{1}{3} & -\frac{4}{9} \\ 3 & 0 & 0 \\ -7 & \frac{2}{3} & -\frac{1}{9} \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 3 \\ -1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & \frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{11}{3} & \frac{4}{3} \\ -2 & 1 & -1 \end{bmatrix}$$
$$= \begin{bmatrix} -\frac{13}{3} & \frac{17}{3} & -\frac{2}{3} \\ -\frac{19}{3} & 7 & -\frac{5}{3} \\ 0 & -\frac{1}{3} & 0 \end{bmatrix}.$$

El lector puede verificar que $A = PBP^{-1}$.

Teorema 2.35 *Todo vector en $\mathbb{R}^n$ es una combinación lineal de vectores de los espacios propios generalizados $\mathcal{V}_j$.*

Demostración.

Supóngase que $\mathbb{V}$ es el subespacio de $\mathbb{R}^n$ formado por los vectores de la forma $\vec{u}_1 + \ldots + \vec{u}_r$, donde $\vec{u}_j \in \mathcal{V}_j$. Se necesita probar que $\mathbb{V} = \mathbb{R}^n$. Supongamos que $\mathbb{V}$ es un subespacio adecuado. Entonces, se escoge una base $\{\vec{v}_1 + \ldots + \vec{v}_s\}$ de $\mathbb{V}$ y se extiende este conjunto a una base $\mathcal{B}$ de $\mathbb{R}^n$. En esta base la matriz $[A]_{\mathcal{B}}$ está particionada como sigue:

$$[A]_{\mathcal{B}} = \begin{bmatrix} A_{11} & A_{12} \\ \mathrm{O} & A_{22} \end{bmatrix},$$

donde A_{22} es una matriz de tamaño $(n-s)\times(n-s)$. Los valores propios de A_{22} son valores propios de A. Como todos los valores propios distintos y vectores propios de A son considerados en $\mathbb{V}$ (es decir, en A_{11}), se tiene una contradicción. Por tanto, $\mathbb{V} = \mathbb{R}^n$, como se quería demostrar. ■

Teorema 2.36 *Sea $\mathfrak{V}_\mathrm{j}$ una base para el espacio propio generalizado $\mathcal{V}_j$ y sea $\mathfrak{V}$ la unión de los conjuntos $\mathfrak{V}_\mathrm{j}$. Entonces, $\mathfrak{V}$ es una base para $\mathbb{R}^n$.*

Demostración.

Vamos a probar que los vectores en $\mathfrak{V}$ generan $\mathbb{R}^n$. Por el Teorema 2.35 se tiene que todo vector en $\mathbb{R}^n$ es una combinación lineal de vectores en $\mathcal{V}_j$, pero cada vector en $\mathcal{V}_j$ es una combinación lineal de vectores en $\mathfrak{V}_\mathrm{j}$. Por lo tanto, los vectores en $\mathfrak{V}$ generan $\mathbb{R}^n$.

Ahora, demostremos que los vectores en $\mathfrak{V}$ son linealmente independientes. Supongamos que una combinación lineal de vectores en $\mathfrak{V}$ suma $\vec{0}$. Es posible escribir esta suma como

$$\vec{v}_1 + \ldots + \vec{v}_r = \vec{0},$$

donde $\vec{v}_j$ es la combinación lineal de vectores en $\mathfrak{V}_\mathrm{j}$. El Teorema 2.34 indica que cada $\vec{v}_j = \vec{0}$. Como $\mathfrak{V}_\mathrm{j}$ es una base para $\mathcal{V}_j$, se deduce que los coeficientes de las combinaciones lineales $\vec{v}_j$ deben ser todos cero. Por lo tanto, los vectores en $\mathfrak{V}$ son linealmente independientes.

Por el Teorema 1.27, se deduce que $\mathfrak{V}$ es una base de $\mathbb{R}^n$. ■

Ejercicios 2.5

1. Encuentre la matriz P que satisface que $A = PBP^{-1}$ para

$$A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 42 & 130 & 25 \\ -8 & -24 & -5 \\ -23 & -73 & -13 \end{bmatrix}.$$

2. Determine los vectores propios generalizados para las matrices

$$a. \begin{bmatrix} -12 & 7 \\ -7 & 2 \end{bmatrix}. \qquad b. \begin{bmatrix} -10 & -7 \\ 7 & 4 \end{bmatrix}. \qquad c. \begin{bmatrix} 4 & -1 \\ 1 & 2 \end{bmatrix}.$$

$$d. \begin{bmatrix} 5 & -3 & -2 \\ 8 & -5 & -4 \\ -4 & 3 & 3 \end{bmatrix}. \qquad e. \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ -2 & -3 & -2 \end{bmatrix}.$$

Capítulo 3

Descomposición de matrices

Una factorización de una matriz A es una ecuación que expresa a A como un producto de dos o más matrices. Por ejemplo, la descomposición de Sylvester y los teoremas de diagonalización dados en 2.19 y 2.25 son algunos casos de descomposición de una matriz. Estas descomposiciones son de interés especial cuando algunos de los factores son matrices ortogonales; la razón es que las transformaciones ortogonales preservan normas y ángulos. Infortunadamente, como sabemos, no todas las matrices pueden ser factorizadas como $A = PDP^{-1}$ con D diagonal. Sin embargo, para cualquier matriz A es posible obtener una de las factorizaciones que se presentan en este capítulo, las cuales son importantes desde el punto de vista teórico, práctico y numérico.

3.1 Triangularización de una matriz

En esta sección nos centraremos en el estudio de diversas factorizaciones de una matriz A como producto de matrices triangulares.

Teorema 3.1 *Sea A una matriz real de tamaño $m \times n$, entonces existe una matriz L triangular inferior no singular de tamaño $m \times m$ tal que*

$$A = LS, \tag{3.1}$$

donde S es la matriz escalonada de A de tamaño $m \times n$, obtenida sin intercambio de filas.

Demostración.

Sea $A = [a_{kl}]$, empecemos con la primera fila no nula de A. Supongamos, sin pérdida de generalidad, que la primera fila de A es no nula. Sea a_{1j} el primer elemento no nulo en la primera fila. Tomemos cualquier a_{ij} con $2 \leq i \leq m$; si $a_{ij} = 0$, no se hace nada; si $a_{ij} \neq 0$, se multiplica la primera fila por $-a_{ij}/a_{1j}$ y se suma a la i-ésima fila. Esta operación hace los elementos (i, j) cero. Dicha operación es equivalente a premultiplicar a A por la matriz elemental $E_{i1}\big(-a_{ij}/a_{1j}\big)$, la cual es una matriz triangular inferior. Así, hemos usado el elemento $(1, j)$, es decir a_{1j}, como un pivote para eliminar todos los otros elementos de la j-ésima columna. La matriz resultante o matriz reducida es obtenida matemáticamente premultiplicando a A sucesivamente por un número finito de matrices triangulares inferiores cuyo producto es también una matriz triangular inferior. Ahora continuemos con la matriz reducida, tomemos la segunda fila. Si todos sus elementos son iguales a cero, se pasa a la tercera fila. Si no se encuentra cualquier vector fila no nulo entre la segunda, tercera, $\ldots, m$-ésima fila, el proceso termina. La matriz reducida es claramente una forma escalonada. En otro caso, localice el primer vector no nulo entre las $m - 1$ filas de la matriz reducida empezando desde la segunda. Repita el proceso de eliminar todas las entradas debajo del primer elemento no nulo (pivote) del vector escogido no nulo. Repita este proceso hasta que no pueda encontrar ningún otro vector no nulo en la matriz reducida. La matriz reducida es claramente una forma escalonada. La matriz S es simplemente el producto de todas las matrices triangulares inferiores empleadas durante el proceso. Claramente, S es una matriz triangular inferior no singular. Esto completa la prueba. ■

Definición 3.1 Descomposición LS

Una factorización como la indicada en (3.1)*, es decir, como el producto de una matriz no singular triangular inferior L y una forma escalonada S, si existe, se llama descomposición LS de la matriz A.*

Ejemplo 3.1 *Hallar una factorización LS de la matriz:*

$$A = \begin{bmatrix} 2 & 4 & -2 & 0 \\ 1 & 1 & -3 & 1 \\ 2 & 3 & -4 & 4 \end{bmatrix}.$$

Solución.

Se procede en dos columnas como sigue:

Reducción de A a S	Creación de L a partir de I
$A = \begin{bmatrix} 2 & 4 & -2 & 0 \\ 1 & 1 & -3 & 1 \\ 2 & 3 & -4 & 4 \end{bmatrix}$	$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Dividir la fila 1 por 2	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 1 & -3 & 1 \\ 2 & 3 & -4 & 4 \end{bmatrix}$	$\sim \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Sumar la fila 1 multiplicada por -1 a la fila 2	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & -1 & -2 & 1 \\ 2 & 3 & -4 & 4 \end{bmatrix}$	$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Sumar la fila 1 multiplicada por -2 a la fila 3	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & -1 & -2 & 1 \\ 0 & -1 & -2 & 4 \end{bmatrix}$	$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$
Multiplicar la fila 2 por -1	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & -1 & -2 & 4 \end{bmatrix}$	$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$
Sumar una vez la fila 2 a la fila 3	
$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$	$\sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & -1 & 1 \end{bmatrix}$

Dividir la fila 3 por 3

$$\sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = S \qquad \sim \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & -1 & 3 \end{bmatrix} = L.$$

El lector puede verificar que $A = LS$.

Teorema 3.2 *Sea A una matriz real de tamaño $m \times n$ cuya forma escalonada puede hallarse sin intercambio de filas. Entonces, A tiene una única factorización LS si y solo si $\rho(A) = m$.*

Demostración.

Supongamos que $\rho(A) = m$ y que $A = LS$ y $A = MV$, siendo L y M matrices triangulares inferiores y S y V matrices en forma escalonada.

Multiplicando por la izquierda, la igualdad

$$LS = MV$$

por M^{-1}, se obtiene que

$$M^{-1}LS = V. \tag{3.2}$$

Sea $N = M^{-1}L$. Si se logra probar que $N = I$, se tendrá que $L = M$ y sustituyendo en (3.2), se obtendrá $S = V$.

Se procede por inducción sobre m. Si $m = 1$, entonces

$$A = \begin{bmatrix} 0 & \dots & a_{1,r+1} & \dots & a_{1,n} \end{bmatrix},$$

y la única factorización posible es

$$A = [a_{1,r+1}] \begin{bmatrix} 0 & \dots & 1 & \dots & \frac{a_{1,n}}{a_{1,r+1}} \end{bmatrix}.$$

Supongamos que el resultado es cierto para matrices con $m-1$ filas. Si la primera columna $\vec{s}_1$ de S fuera nula, la primera columna $\vec{v}_1$ de V también lo sería, ya que

$$\vec{v}_1 = N\vec{s}_1 = N\vec{0} = \vec{0}$$

y, recíprocamente, por ser N no singular. Se puede por lo tanto suponer, sin pérdida de generalidad, que las primeras columnas de S y V son no

nulas. Entonces la primera columna tiene un uno principal en la primera fila.

Nótese, además, que N es triangular inferior por ser producto de matrices triangulares inferiores. Si se reescribe el producto $NS = V$ particionando N en submatrices $\big[1+(m-1)\big] \times \big[1+(m-1)\big]$ y las matrices S y V en submatrices $\big[1+(m-1)\big] \times \big[1+(n-1)\big]$, se tiene que

$$NS = \begin{bmatrix} a & \mathrm{O} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} 1 & S_{12} \\ \mathrm{O} & S_{22} \end{bmatrix} = \begin{bmatrix} 1 & V_{12} \\ \mathrm{O} & V_{22} \end{bmatrix},$$

de donde

$$\begin{aligned} a &= 1, & aS_{12} &= V_{12}, \\ N_{21} &= \mathrm{O}, & N_{22}S_{22} &= V_{22}. \end{aligned}$$

Ahora bien, N_{22} es una submatriz real de tamaño $(m-1) \times (m-1)$ que es triangular inferior y no singular, S_{22} y V_{22} son submatrices en forma escalonada. Por la hipótesis de inducción,

$$N_{22} = I_{m-1},$$

y por lo tanto

$$N = \begin{bmatrix} 1 & \mathrm{O} \\ \mathrm{O} & I_{m-1} \end{bmatrix} = I_m.$$

Si se supone que $r = \rho(A) < m$, aplicando el algoritmo se obtiene la factorización

$$A = \begin{bmatrix} L_{11} & \mathrm{O} \\ L_{21} & I_{m-r} \end{bmatrix} \begin{bmatrix} S_{11} \\ \mathrm{O} \end{bmatrix}.$$

Pero es obvio que si L_{22} es una submatriz triangular inferior no singular de tamaño $\big(m-r\big) \times \big(m-r\big)$ cualquiera, también se puede escribir

$$A = \begin{bmatrix} L_{11} & \mathrm{O} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} S_{11} \\ \mathrm{O} \end{bmatrix},$$

y la factorización no es única. ■

Corolario 3.2.1 *Sea A una matriz no singular de tamaño $n \times n$ cuya forma escalonada puede hallarse sin intercambio de filas. Entonces A tiene una única factorización LS.*

Demostración.

Supongamos que $A = L_1S_1$ y $A = L_2S_2$ son dos de dichas factorizaciones. Nótese que tanto L_1^{-1} como L_2^{-1} también son triangulares inferiores y S_1^{-1} y S_2^{-1} son triangulares superiores, que además tienen unos en la diagonal principal por ser A no singular.

Ahora bien, de $L_1S_1 = L_2S_2$ se obtiene que

$$L_2^{-1}L_1 = S_2S_1^{-1}.$$

Es claro que $L_2^{-1}L_1$ es triangular inferior por ser producto de triangulares inferiores y que $S_2S_1^{-1}$ es triangular superior por ser producto de triangulares superiores. Como son iguales, se concluye que el producto debe ser diagonal. Además, S_2 y S_1^{-1} tienen unos en la diagonal principal y, por lo tanto, $S_2S_1^{-1}$ también tiene unos en la diagonal principal.

En definitiva $L_2^{-1}L_1 = S_2S_1^{-1} = I_n$, de donde se deduce la unicidad. ■

Teorema 3.3 *Si no ocurren intercambios de filas durante la reducción de una matriz A de tamaño $m \times n$ a una matriz escalonada S, entonces A puede ser factorizada como*

$$A = LDU, \tag{3.3}$$

en donde L es triangular inferior de tamaño $m \times m$, D es la matriz de pivotes de tamaño $m \times m$ y U es una matriz escalonada.

Demostración.

Sea

$$S = E_kE_{k-1} \dots E_2E_1A$$

la forma escalonada de A, donde E_i, $i = 1, 2, \dots, k$ son matrices elementales de tipo R_1 y R_2, puesto que no se han realizado intercambio de filas. Entonces resulta que

$$A = E_1^{-1}E_2^{-1} \dots E_{k-1}^{-1}E_k^{-1}S = LS.$$

Después de que se determina $A = LS$, se continúa factorizando S como

$$S = DD^{-1}S = DU,$$

en donde D es la matriz diagonal de pivotes cuyo elemento diagonal en la p-ésima fila es 1 si la p-ésima fila de S es $\vec{0}$, y es a si a es el primer elemento no nulo de la p-ésima fila de S. La matriz escalonada $U = D^{-1}S$. Entonces, se puede reescribir el producto $A = LS$ como

$$A = LDU = MU,$$

en donde $M = LD$. Esto prueba el teorema. ■

Ejemplo 3.2 *Hallar la factorización LDU de la matriz dada en el Ejemplo 3.1.*

Solución.

Del Ejemplo 3.1, se tiene que

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Los pivotes eran 2, -1 y 3. Luego, su factorización LDU es

$$\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Teorema 3.4 Factorización única

Sea A una matriz no singular de tamaño $n \times n$. Una factorización de la forma (3.3) *está determinada de manera única si y solo si*

1. $L = [l_{ij}]$ *es triangular inferior con* $l_{ii} = 1$ *para* $i = 1, 2, \ldots, n$,

2. $U = [u_{ij}]$ *es triangular superior con* $u_{ii} = 1$ *para* $i = 1, 2, \ldots, n$,

3. $D = [d_{ij}]$ *es diagonal con* $d_{ii} \neq 0$ *para* $i = 1, 2, \ldots, n$.

Demostración.

Supongamos que $A = L_1 D_1 U_1$ y $A = L_2 D_2 U_2$ son dos factorizaciones de A distintas. Nótese que tanto L_1^{-1} como L_2^{-1} también son triangulares inferiores y U_1^{-1} y U_2^{-1} son triangulares superiores, que además tienen unos en la diagonal principal por ser A no singular.

Ahora bien, de $L_1D_1U_1 = L_2D_2U_2$ se obtiene que

$$U_1U_2^{-1} = D_1^{-1}L_1^{-1}L_2D_2.$$

El lado izquierdo es un producto de dos matrices triangulares superiores con elementos en la diagonal principal iguales a uno. Dicho producto debe ser otra matriz del mismo tipo. Por otra parte, el lado derecho es una matriz triangular inferior. Esto obliga a que ambos lados sean precisamente la matriz identidad: la única matriz que al mismo tiempo es triangular superior con diagonal unitaria y también triangular superior. Así, $U_1U_2^{-1} = I_n$, y después de multiplicar por U_2 se tiene que $U_1 = U_2$.

Análogamente, $L_1 = L_2$ y, finalmente, $D_1 = D_2$. ■

Teorema 3.5 *Si A es una matriz simétrica que se puede factorizar como $A = LDU$ sin intercambios de filas que destruyan la simetría, entonces la triangular superior U es la transpuesta de la triangular inferior L. En otras palabras, toda matriz simétrica tiene una factorización simétrica de la forma $A = LDL^t$.*

Demostración.

Supongamos que A puede factorizarse como $A = LDU$; tomando la transpuesta, se tiene que

$$A^t = \big(LDU\big)^t = U^tD^tL^t = U^tDL^t.$$

Como A es simétrica, es igual a A^t; así resulta que tenemos dos factorizaciones de A en triangular inferior por diagonal por triangular superior. (L^t es triangular superior con unos en la diagonal, exactamente como U). De acuerdo con el Teorema 3.4, esta factorización es única; por lo tanto, L^t debe ser idéntica a U, lo cual completa la prueba. ■

Ejemplo 3.3 *Hallar una factorización LDU para la matriz:*

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 3 & 12 & 18 \\ 5 & 18 & 30 \end{bmatrix}.$$

Solución.

Se procede en dos columnas como en el Ejemplo 3.1.

Reducción de A a S	Creación de L a partir de I
$A = \begin{bmatrix} 1 & 3 & 5 \\ 3 & 12 & 18 \\ 5 & 18 & 30 \end{bmatrix}$	$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Sumar la fila 1 multiplicada por -3 a la fila 2	
$\sim \begin{bmatrix} 1 & 3 & 5 \\ 0 & 3 & 3 \\ 5 & 18 & 30 \end{bmatrix}$	$\sim \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Sumar la fila 1 multiplicada por -5 a la fila 3	
$\sim \begin{bmatrix} 1 & 3 & 5 \\ 0 & 3 & 3 \\ 0 & 3 & 5 \end{bmatrix}$	$\sim \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 0 & 1 \end{bmatrix}$
Sumar la fila 2 multiplicada por -1 a la fila 3	
$\sim \begin{bmatrix} 1 & 3 & 5 \\ 0 & 3 & 3 \\ 0 & 0 & 2 \end{bmatrix} = S$	$\sim \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 1 & 1 \end{bmatrix} = L.$

Por tanto, se tiene que $A = LS$, y factorizando a S, se llega a

$$S = \begin{bmatrix} 1 & 3 & 5 \\ 0 & 3 & 3 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Nótese que $L^t = U$.

Teorema 3.6 Descomposición triangular LU

Sea A una matriz real de tamaño $n \times n$ tal que todos sus menores principales son no nulos. Entonces A puede ser factorizada como

$$A = LU, \tag{3.4}$$

donde L es una matriz triangular inferior y U es una matriz no singular triangular superior, cada una de tamaño $n \times n$.

Demostración.

Queda como ejercicio para el lector. ■

Definición 3.2 Matriz triangularizable

Una matriz A de tamaño $n \times n$ es triangularizable si existe una matriz no singular P tal que

$$P^{-1}AP = T \tag{3.5}$$

es triangular.

Teorema 3.7 Descomposición de Schur

Si A es una matriz real de tamaño $n \times n$ con valores propios reales $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces existe una matriz ortogonal Q tal que

$$Q^t AQ = T = [t_{ij}], \tag{3.6}$$

donde T es una matriz triangular superior con elementos en la diagonal $t_{ii} = \lambda_i$, $i = 1, 2, \ldots, n$. (Es decir: toda matriz cuadrada real con valores propios reales es ortogonalmente semejante a una matriz triangular superior).

Demostración.

La demostración es por inducción sobre n. Si $n = 1$, A es una matriz real de tamaño 1×1 que es triangular. La matriz ortogonal es $Q = [1]$.

Supongamos que toda matriz real de tamaño $(n-1) \times (n-1)$ es triangularizable por una matriz ortogonal. Sea A una matriz real de tamaño $n \times n$ con valores propios reales. Sea $\vec{v}_1 \in \mathbb{R}^n$ un vector propio unitario asociado al valor propio λ_1. Denotemos por $\mathbb{W}$ el complemento ortogonal a $\vec{v}_1$ de dimensión $n-1$. Sea $\{\vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n\}$ una base ortonormal de $\mathbb{W}$. Luego, cada vector $\vec{X}$ de $\mathbb{W}$ tiene la forma

$$\vec{X} = a_2\ \vec{v}_2 + a_3\ \vec{v}_3 + \ldots + a_n\ \vec{v}_n.$$

La matriz de cambio de base de la *base canónica* de $\mathbb{R}^n$ a la base $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ es la matriz S cuyas columnas son los elementos de los vectores $\vec{v}_i$. Luego,

$$\begin{aligned} AS &= \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}. \end{aligned}$$

Por lo tanto,

$$S^{-1}AS = S^{-1}\begin{bmatrix} \lambda_1\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}.$$

Pero como S es ortogonal, se tiene que $S^{-1} = S^t$. Por consiguiente

$$S^tAS = \begin{bmatrix} \lambda_1 & x_{12} & \ldots & x_{1n} \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{bmatrix},$$

donde $x_{1i} = \vec{v}_1^{\,t}A\vec{v}_i$ y A_1 es una matriz real de tamaño $(n-1)\times(n-1)$.

La prueba se completa por inducción: sea R_1 una matriz ortogonal de tamaño $\big(n-1\big)\times\big(n-1\big)$ tal que $(R_1)^tA_1R_1 = T_1$, con T_1 triangular superior, por la hipótesis de inducción. Entonces, la matriz

$$R = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & R_1 & \\ 0 & & & \end{bmatrix}$$

también es una matriz ortogonal y

$$\begin{aligned} (SR)^tA(SR) = R^t\big(S^tAS\big)R &= \begin{bmatrix} 1 & \vec{0}^t \\ \vec{0} & R_1^t \end{bmatrix}\begin{bmatrix} \lambda_1 & \vec{x}^t \\ \vec{0} & A_1^t \end{bmatrix}\begin{bmatrix} 1 & \vec{0}^t \\ \vec{0} & R_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \vec{x}^tR_1 \\ \vec{0} & T_1 \end{bmatrix}, \end{aligned}$$

donde $\vec{x}^t = [x_{12}\ \ x_{13}\ \ \ldots\ \ x_{1n}]$. La matriz $SR = Q$ es el producto de dos matrices ortogonales, por lo tanto es también una matriz ortogonal. Así, Q^tAQ es una matriz triangular superior y nuestra prueba queda completa. ■

Ejemplo 3.4 *Dada la matriz:*

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{bmatrix}.$$

Encuentre la triangularización de A.

Solución.

El polinomio característico de A es

$$p_A(\lambda) = (2-\lambda)^3.$$

Entonces, el único valor propio es $\lambda = 2$ (de multiplicidad algebraica 3). Los vectores propios correspondientes a $\lambda = 2$ son $\vec{u}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ y $\vec{u}_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$. La idea básica del teorema de Schur consiste en construir una base de $\mathbb{R}^3$ con el mayor número posible de vectores propios.

Si tomamos por ejemplo $\vec{u}_3 = \vec{u}_1 \times \vec{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$, el conjunto $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ es una base de $\mathbb{R}^3$. Mediante el algoritmo de Gram-Schmidt, obtenemos la base ortonormal $\left\{\vec{v}_1, \vec{v}_2, \vec{v}_3\right\}$, donde

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

La matriz ortogonal Q es

$$Q = \begin{bmatrix} 0 & \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} \\ 0 & \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} \\ 1 & 0 & 0 \end{bmatrix}.$$

La matriz triangular es

$$T = Q^t A Q = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} & 0 \\ -\frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} \\ 0 & \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} \\ 1 & 0 & 0 \end{bmatrix},$$

y al realizar las operaciones, se llega a

$$T = \begin{bmatrix} 2 & 0 & \sqrt{5} \\ 0 & 2 & 5 \\ 0 & 0 & 2 \end{bmatrix}.$$

Obsérvese que los elementos de la diagonal principal de la matriz T son los valores propios de la matriz A.

Definición 3.3 Triangularización simultánea

Si P es una matriz no singular tal que $P^{-1}AP$ y $P^{-1}BP$ son ambas diagonales (triangulares), entonces se dice que A y B son diagonalizables (triangularizables) simultáneamente.

Teorema 3.8 *Si A y B son matrices de tamaño $n \times n$ y $AB = BA$, entonces A y B tienen un vector propio en común.*

Demostración.

Sea λ_k un valor propio de A y sea $\mathcal{B}_k = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ una base del $\ker\{A - \lambda_k I\}$. Entonces, cualquier combinación lineal no nula de $\mathcal{B}_k$ constituye un vector propio correspondiente a λ_k y, recíprocamente, cualquier vector propio de A correspondiente a λ_k está en $\text{gen}\{\mathcal{B}_k\}$.

Si para algún $j = 1, 2, \ldots, k$, $B\vec{v}_j = \vec{0}$, entonces $\vec{v}_j$ es un vector propio de B asociado al valor propio 0 y el teorema queda demostrado. Suponga que $B\vec{v}_j \neq \vec{0}$, entonces

$$\begin{aligned} A(B\vec{v}_j) &= (BA)\vec{v}_j && \text{puesto que } AB = BA, \\ &= B(\lambda_k \vec{v}_j) && \text{puesto que } \vec{v}_j \text{ es un vector propio de } A \\ &= \lambda_k (B\vec{v}_j). \end{aligned}$$

Luego

$$B\vec{v}_j \in \text{gen}\{\mathcal{B}_k\} \qquad \text{para} \qquad j = 1, 2, \ldots, k.$$

Al expresar cada $B\vec{v}_j$ como una combinación lineal de $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$, se tiene

$$B\vec{v}_j = \sum_{i=1}^{k} c_{ij}\vec{v}_i, \qquad j = 1, 2, \ldots, k.$$

Este sistema de ecuaciones escrito en forma matricial queda

$$B\underbrace{\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_k \end{bmatrix}}_{B \qquad V} = \underbrace{\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_k \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1k} \\ c_{21} & c_{22} & \dots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \dots & c_{kk} \end{bmatrix}}_{C},$$

donde $C = [c_{ij}]$ sea μ_1 un valor propio de C y $\vec{u}^{\,t} = [\alpha_1 \;\; \alpha_2 \;\; \dots \;\; \alpha_k]$ el vector propio asociado a μ_1. Entonces

$$\begin{aligned} \big(BV\big)\vec{u} &= (VC)\vec{u} \\ &= V\big(\mu_1\vec{u}\big) \qquad \text{puesto que } \vec{u} \text{ es un vector propio de } C \\ &= \mu_1\big(V\vec{u}\big). \end{aligned}$$

Luego, $V\vec{u}$ es un vector propio de B, pero nótese que

$$V\vec{u} = \sum_{i=1}^{k} \alpha_i\vec{v}_i \in \text{gen}\big\{\mathcal{B}_k\big\},$$

y por consiguiente $V\vec{u}$ es un vector propio de A. ■

Ejemplo 3.5 *Sean*

$$A = \begin{bmatrix} 2 & -1 & -1 \\ 2 & -1 & -2 \\ -1 & 1 & 2 \end{bmatrix} \qquad y \qquad B = \begin{bmatrix} 7 & -1 & -1 \\ 11 & 4 & 7 \\ -10 & 1 & -2 \end{bmatrix}.$$

Determine si $AB = BA$ y encuentre un vector propio que sea común a las matrices A y B.

Solución.

Al realizar el producto entre A y B, se obtiene

$$AB = \begin{bmatrix} 13 & -7 & -7 \\ 23 & -8 & -5 \\ -16 & 7 & 4 \end{bmatrix} = BA.$$

Para encontrar el vector propio común a A y B, primero se establece la ecuación característica de la matriz A,

$$p_A(\lambda) = -\lambda^3 + 3\lambda^2 - 3\lambda + 1 = -(\lambda - 1)^3 = 0,$$

luego $\lambda = 1$ es el único valor propio con multiplicidad algebraica igual a tres. Entonces,

$$(A - \lambda I)\vec{v} = (A - I)\vec{v} = \begin{bmatrix} 1 & -1 & -1 \\ 2 & -2 & -2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Esto conduce a $x - y = z$. Luego, los vectores propios linealmente independientes son $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$.

Ahora, se calcula

$$B\vec{v}_1 = \begin{bmatrix} 6 \\ 18 \\ -12 \end{bmatrix} \qquad \text{y} \qquad B\vec{v}_2 = \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix}.$$

Al expresar $B\vec{v}_i$ como una combinación lineal de $\{\vec{v}_1, \vec{v}_2\}$, se tiene

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \qquad \text{y} \qquad C = \begin{bmatrix} 6 & 0 \\ 18 & -3 \end{bmatrix}.$$

Nótese que $BV = VC$. Por otra parte, los valores propios de C son $\mu_1 = 6$ y $\mu_2 = -3$, y los vectores propios correspondientes son $\vec{u}_1^t = [1 \;\; 2]$ y $\vec{u}_2^t = [0 \;\; 1]$, respectivamente. Entonces, $V\vec{u}_1 = [1 \;\; 2 \;\; -1]^t$ y $V\vec{u}_2 = [0 \;\; 1 \;\; -1]^t$ son vectores propios comunes a las matrices A y B.

Teorema 3.9 *Si A y B son matrices de tamaño $n \times n$ y $AB = BA$, entonces A y B pueden ser triangularizadas ortogonalmente en forma simultánea.*

Demostración.

La demostración es por inducción sobre n. Por el Teorema 3.8, existe $\vec{v}_1 \in \mathbb{R}^n$ normalizado tal que

$$A\vec{v}_1 = \lambda_1\vec{v}_1 \qquad \text{y} \qquad B\vec{v}_1 = \mu_1\vec{v}_1.$$

De manera análoga, como en la demostración del Teorema 3.7 se construye una matriz S ortogonal cuya primera columna sea $\vec{v}_1$, entonces

$$S^tAS = \begin{bmatrix} \lambda_1 & x_{12} \ \dots \ x_{1n} \\ 0 & \\ \vdots & \quad A_1 \quad \\ 0 & \end{bmatrix} \quad \text{y} \quad S^tBS = \begin{bmatrix} \mu_1 & y_{12} \ \dots \ y_{1n} \\ 0 & \\ \vdots & \quad B_1 \quad \\ 0 & \end{bmatrix},$$

donde A_1 y B_1 son matrices reales de tamaño $\big(n-1\big)\times\big(n-1\big)$. Por otra parte,

$$\begin{aligned} \big(S^tAS\big)\big(S^tBS\big) &= S^t(AB)S && \text{puesto que } S^t = S^{-1}, \\ &= S^t(BA)S && \text{puesto que } AB = BA, \\ &= \big(S^tBS\big)\big(S^tAS\big), \end{aligned}$$

y por la multiplicación por bloques, se tiene

$$\begin{bmatrix} \lambda_1\mu_1 & \vdots & \lambda_1\vec{y}^{\,t} + \vec{x}^{\,t}B_1 \\ \cdots & . & \cdots\cdots \\ \vec{0} & \vdots & A_1B_1 \end{bmatrix} = \begin{bmatrix} \lambda_1\mu_1 & \vdots & \mu_1\vec{x}^{\,t} + \vec{y}^{\,t}A_1 \\ \cdots & . & \cdots\cdots \\ \vec{0} & \vdots & B_1A_1 \end{bmatrix},$$

donde $\vec{x}^{\,t} = [x_{12} \ \ x_{13} \ \ \dots \ \ x_{1n}]$, $\vec{y}^{\,t} = [y_{12} \ \ y_{13} \ \ \dots \ \ y_{1n}]$, y por lo tanto $A_1B_1 = B_1A_1$. Los detalles para culminar la prueba son análogos a los del Teorema 3.7 y se dejan como ejercicio para el lector. ■

Teorema 3.10 *Si A y B son matrices simétricas de tamaños $n\times n$, entonces $AB = BA$ si y solo si existe una matriz ortogonal Q de tamaño $n\times n$ tal que Q^tAQ y Q^tBQ sean ambas diagonales.*

Demostración.

Supongamos que $AB = BA$. Por el Teorema 3.9, existe una matriz ortogonal Q tal que Q^tAQ y Q^tBQ son ambas triangulares. Pero como A y B son simétricas, pueden ser diagonalizables ortogonalmente. De aquí que Q^tAQ y Q^tBQ sean diagonales. Por otra parte, si $Q^tAQ = D_A$ y $Q^tBQ = D_B$ son ambas diagonales, entonces

$$\begin{aligned} AB &= \big(QD_AQ^t\big)\big(QD_BQ^t\big) = QD_AD_BQ^t \\ &= QD_BD_AQ^t = \big(QD_BQ^t\big)\big(QD_AQ^t\big) \\ &= BA. \end{aligned}$$

Esto completa la demostración. ■

3.2 Factorización QR

Esta factorización se usa ampliamente en los programas de computadora para resolver sistemas lineales, para determinar aproximaciones por mínimos cuadrados y para determinar los valores propios de una matriz.

Teorema 3.11 Factorización QR

Sea A una matriz real de tamaño $m \times n$ con $\rho(A) = n$. Entonces existen una matriz Q de tamaño $m \times n$ cuyas columnas son ortonormales y una matriz no singular R de tamaño $n \times n$ triangular superior tales que

$$A = QR. \tag{3.7}$$

Demostración.

Como $\rho(A) = n$, entonces sus columnas son linealmente independientes. Sean $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ las columnas de A, las cuales constituyen una base para el espacio generado por las columnas de A $\left[Col(A)\right]$ y mediante el proceso de Gram-Schmidt, se ortonormalizan estos vectores para obtener $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ una base ortonormal para el espacio generado por las columnas de A. Sea

$$Q = \left[\begin{array}{cccc} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{array}\right].$$

Al expresar cada $\vec{x}_j$ como una combinación lineal de $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, se tiene

$$\vec{x}_j = \sum_{i=1}^{n} r_{ij}\vec{v}_i, \qquad j = 1, 2, \ldots, n.$$

Este sistema de ecuaciones escrito en forma matricial queda

$$\underbrace{\left[\begin{array}{cccc} \vec{x}_1 & \vec{x}_2 & \ldots & \vec{x}_n \end{array}\right]}_{A} = \underbrace{\left[\begin{array}{cccc} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{array}\right]}_{Q} \underbrace{\left[\begin{array}{cccc} r_{11} & r_{12} & \ldots & r_{1n} \\ r_{21} & r_{22} & \ldots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \ldots & r_{nn} \end{array}\right]}_{\left[\begin{array}{cccc} \vec{r}_1 & \vec{r}_2 & \ldots & \vec{r}_n \end{array}\right]},$$

donde

$$\vec{r}_k = \begin{bmatrix} r_{1k} \\ r_{2k} \\ \vdots \\ r_{nk} \end{bmatrix}, \qquad k = 1, 2, \ldots, n.$$

Por otra parte, como $\vec{v}_j$ es ortogonal a la base gen$\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ si $j > k$, entonces es ortogonal a $\vec{x}_k$. Por lo tanto, $r_{jk} = 0$ para $j > k$, ya que

$$r_{jk} = \vec{v}_j{}^t \vec{x}_k = \vec{v}_j \cdot \vec{x}_k.$$

Sea $R = \begin{bmatrix} \vec{r}_1 & \vec{r}_2 & \ldots & \vec{r}_n \end{bmatrix}$, entonces

$$A = QR = Q \begin{bmatrix} r_{11} & r_{12} & r_{13} & \ldots & r_{1n} \\ 0 & r_{22} & r_{23} & \ldots & r_{2n} \\ 0 & 0 & r_{33} & \ldots & r_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & r_{nn} \end{bmatrix}.$$

Ahora, mostremos que R es no singular. Consideremos el sistema lineal $R\vec{b} = \vec{0}$ y multipliquemos por Q a la izquierda, es decir

$$\begin{array}{cccc} \underbrace{QR}\, & \vec{b} & = & \underbrace{Q\vec{0}} \\ A & \vec{b} & = & \vec{0}. \end{array}$$

Pero como las columnas de A son linealmente independientes, el sistema homogéneo $A\vec{b} = \vec{0}$ solo tiene la solución trivial. Por lo tanto, R es no singular. ■

Nota 3.1 Para el caso en que A sea una matriz real de tamaño $m \times n$ con $\rho(A) = m$, entonces se puede encontrar una factorización de manera análoga a (3.7) de la forma

$$A = LQ, \tag{3.8}$$

donde L es una matriz real de tamaño $m \times m$ triangular inferior y no singular, y Q es una matriz real de tamaño $m \times n$ cuyas filas son ortonormales.

Ejemplo 3.6 *Encuentre una factorización QR de*

$$A = \begin{bmatrix} 4 & 25 \\ 0 & 0 \\ 3 & -25 \end{bmatrix}.$$

Solución.

Denotemos las columnas de A por

$$\vec{x}_1 = \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix} \qquad \text{y} \qquad \vec{x}_2 = \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix}.$$

Si se aplica el algoritmo de Gram-Schmidt al conjunto $\{\vec{x}_1, \vec{x}_2\}$ base para el espacio generado por las columnas de A, como $\|\vec{x}_1\| = 5$, se hace $\vec{v}_1 = \dfrac{\vec{x}_1}{\|\vec{x}_1\|} = \begin{bmatrix} 4/5 \\ 0 \\ 3/5 \end{bmatrix}$. Después,

$$\begin{aligned} \vec{v}_2' &= \vec{x}_2 - (\vec{x}_2 \cdot \vec{v}_1)\vec{v}_1 = \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix} - 5 \begin{bmatrix} 4/5 \\ 0 \\ 3/5 \end{bmatrix} \\ &= \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix} - \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 21 \\ 0 \\ -28 \end{bmatrix}. \end{aligned}$$

Entonces, $\|\vec{v}_2'\| = 35$ y $\vec{v}_2 = \begin{bmatrix} 3/5 \\ 0 \\ -4/5 \end{bmatrix}$. Se puede verificar que $\{\vec{v}_1, \vec{v}_2\}$ es una nueva base ortonormal para el espacio generado por las columnas de A, observando que $\vec{v}_1 \cdot \vec{v}_2 = 0$. Entonces formamos la matriz

$$Q = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 & 3 \\ 0 & 0 \\ 3 & -4 \end{bmatrix}.$$

Para encontrar R, se despeja esta matriz de $A = QR$, de la siguiente manera:

$$Q^t A = Q^t(QR) = IR = R.$$

De esta manera, la matriz R es

$$R = \frac{1}{5}\begin{bmatrix} 4 & 0 & 3 \\ 3 & 0 & -4 \end{bmatrix}\begin{bmatrix} 4 & 25 \\ 0 & 0 \\ 3 & -25 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 0 & 35 \end{bmatrix}.$$

El lector puede verificar que $A = QR$.

Si la matriz A es cuadrada, entonces se puede enunciar el Teorema 3.11 de la siguiente manera.

Teorema 3.12 *Toda matriz cuadrada real A puede expresarse en la forma:*

$$A = QR, \tag{3.9}$$

donde Q es una matriz ortogonal propia y R es triangular superior, con $r_{ii} > 0$, $i = 1, 2, \ldots, n-1$.

Demostración.
Queda como ejercicio para el lector. ■

Corolario 3.12.1 *Si A es ortogonal con $\det A = 1$ entonces, en (3.9), $R = I_n$. Si $\det A = -1$ entonces, los elementos de $R = [r_{ij}]$ cumplen que*

$$r_{ij} = \begin{cases} 1 & si \quad i = j, \\ 0 & si \quad i \neq j, \end{cases}$$

excepto $r_{nn} = -1$.

Demostración.
Si $A = QR$, entonces

$$\det A = \det(QR) = \det R, \qquad \text{ya que} \qquad \det Q = 1$$
$$= r_{nn}\left(\prod_{i=1}^{n-1} r_{ii}\right).$$

Por tanto, se tiene que $r_{nn} = \pm 1$ ya que $\det A = \pm 1$, y la prueba del corolario se completa. ■

Corolario 3.12.2 *Si A es no singular, entonces la representación dada en* (3.9) *es única.*

Demostración.

Supongamos que A es no singular y consideremos dos factorizaciones distintas

$$A = QR \qquad \text{y} \qquad A = Q'R',$$

con Q y Q', ambas ortogonales propias, y R y R' triangulares superiores. Entonces

$$\begin{aligned} I &= Q^tQ'R'R^{-1} = \big(Q^tQ'\big)\big(R'R^{-1}\big) \\ &= \widetilde{Q}\widetilde{R}. \end{aligned}$$

Aquí, la matriz ortogonal I está representada como el producto de una matriz ortogonal propia $\widetilde{Q}$ y una triangular superior $\widetilde{R}$. Por lo tanto, de acuerdo con el Corolario 3.12.1, $\widetilde{R} = I$ y $\widetilde{Q} = I$. Luego, $R' = R$ y $Q' = Q$ de este modo el corolario está probado. ■

Ejercicios 3.1

Para cada una de las siguientes matrices determine (en caso de ser posible) las factorizaciones LU y QR y la descomposición de Schur:

1. $\begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}$. 2. $\begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}$. 3. $\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$. 4. $\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$.

5. $\begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$. 6. $\begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}$. 7. $\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$. 8. $\begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}$.

3.3 Polinomio mínimo

El polinomio característico de una matriz es un instrumento para calcular sus valores propios. En esta sección se estudia el polinomio mínimo de matrices, el cual resulta muy útil para establecer criterios sobre la posibilidad de reducir matrices a formas canónicas simples.

Definición 3.4 Polinomios de matrices

Si A es una matriz real de tamaño $n \times n$, el polinomio $p_n(A)$ denota la matriz que se genera si se reemplaza cada aparición de x en $p_n(x)$ por la matriz A:

$$p_n(A) = a_n A^n + a_{n-1} A^{n-1} + \ldots + a_1 A + a_0 A^0,$$

donde $a_i \in \mathbb{R}$ $\big(i = 0, 1, \ldots, n\big)$ y $A^0 = I_n$.

En consecuencia, se dice que A satisface el polinomio $p_n(x)$ si $p_n(A) = \mathrm{O}$.

Ejemplo 3.7 *Verifique si el polinomio $p_2(x) = x^2 - 3x - 28$ lo satisface la matriz:*

$$A = \begin{bmatrix} 1 & -5 \\ -6 & 2 \end{bmatrix}.$$

Solución.

Si la matriz A satisface dicho polinomio se debe verificar que

$$P_2(A) = A^2 - 3A - 28I = \mathrm{O}.$$

Como

$$A^2 = \begin{bmatrix} 31 & -15 \\ -18 & 34 \end{bmatrix},$$

al sustituir, se tiene que

$$A^2 - 3A - 28I = \begin{bmatrix} 31 & -15 \\ -18 & 34 \end{bmatrix} - 3\begin{bmatrix} 1 & -5 \\ -6 & 2 \end{bmatrix} - 28\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Luego, la matriz A sí satisface el polinomio dado.

Teorema 3.13 Teorema de Cayley-Hamilton

Sean A una matriz de tamaño $n \times n$ y

$$p_A(\lambda) = \det(A - \lambda I_n) = \lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0 = 0$$

su polinomio característico. Entonces, $p_A(A) = \mathrm{O}$. Es decir, A satisface la ecuación

$$A^n + c_{n-1}A^{n-1} + \ldots + c_1 A + c_0 A^0 = \mathrm{O}.$$

El teorema es verdadero para cualquier matriz; sin embargo, la prueba que se presenta a continuación es únicamente para matrices diagonalizables. La demostración para el caso general puede verse en Apostol (1985, p. 249).

Demostración.

Supongamos que A es diagonalizable. Como $p_A(\lambda)$ es una ecuación escalar, al multiplicarla por cualquier vector $\vec{v} \in \mathbb{R}^n$, se tiene

$$\begin{aligned} p_A\big(\lambda\big)\vec{v} &= \big(\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0\big)\vec{v} = \vec{0} \\ &= \lambda^n\vec{v} + c_{n-1}\lambda^{n-1}\vec{v} + \ldots + c_1\lambda\vec{v} + c_0\vec{v} = \vec{0}. \end{aligned}$$

Si $\vec{v}$ es un vector propio correspondiente al valor propio λ, se cumple

$$p_A\big(\lambda\big)\vec{v} = A^n\vec{v} + c_{n-1}A^{n-1}\vec{v} + \ldots + c_1 A\vec{v} + c_0 I_n\vec{v} = \vec{0}. \qquad (3.10)$$

Esto se cumple para todos los vectores propios de A. Pero como A es diagonalizable, tiene n-vectores propios linealmente independientes, luego cualquier otro vector de $\mathbb{R}^n$ puede ser expresado como combinación lineal de estos. Por lo tanto, (3.10) se cumple para todo vector de $\mathbb{R}^n$. De aquí,

$$A^n + c_{n-1}A^{n-1} + \ldots + c_1 A + c_0 I_n = \mathrm{O},$$

es decir que A satisface su propia ecuación característica. ■

Sea A una matriz real de tamaño $n \times n$ y definamos

$$S = \big\{p_n(x) | p_n(A) = \mathrm{O}\big\}, \qquad n \geq 1.$$

Entonces, en S se puede escoger un polinomio no nulo $q(x)$ que tenga grado mínimo; además, se puede suponer que el coeficiente de $q(x)$ correspondiente al término de mayor grado es 1, es decir que $q(x)$ es mónico.

Con estas condiciones, para $q(x)$ se puede demostrar que cualquier otro polinomio $p(x)$ de S es múltiplo de $q(x)$. Esto implica que si en S existiera otro polinomio $r(x)$ mónico del mismo grado de $q(x)$, es decir de grado mínimo, entonces $r(x) = q(x)$.

Definición 3.5 *El polinomio mínimo de una matriz A es el polinomio no nulo de menor grado que es satisfecho por A. Se denotará por $m_A(x)$.*

Como la multiplicidad algebraica de los valores propios de una matriz es a veces distinta de uno, el polinomio característico $p_A(x)$ no es necesariamente el polinomio de grado mínimo satisfecho por A.

Teorema 3.14 *El polinomio mínimo de una matriz cuadrada A es único cuando se impone que el coeficiente del término de mayor exponente (grado) en la indeterminada sea igual a la unidad.*

Demostración.

La prueba se realiza por contradicción. Supongamos que $m_A(x)$ y $m'_A(x)$ son polinomios mínimos de A; por la Definición 3.5, ambos tienen el mismo grado.

Al considerarse los coeficientes dominantes respectivos iguales a la unidad, el polinomio $d(x) = m_A(x) - m'_A(x)$ tiene grado menor que los polinomios mínimos y se anula también para A, y necesariamente $d(x) = 0$. Luego, $m_A(x) = m'_A(x)$. ■

Teorema 3.15 *Si A satisface un polinomio $p_n(x)$, entonces $p_n(x)$ es divisible por $m_A(x)$, polinomio mínimo de A.*

Demostración.

Sea $m_A(x)$ el polinomio mínimo de A, el grado de $p_n(x)$ es mayor que el de $m_A(x)$. Por el algoritmo de la división, sabemos que se pueden encontrar polinomios $q(x)$ y $r(x)$ tales que

$$p_n(x) = m_A(x)q(x) + r(x),$$

con $grad\big[r(x)\big] < grad\big[m_A(x)\big]$. Entonces, $r(x) = p_n(x) - m_A(x)q(x)$, y como $p_n(A) = \mathrm{O}$, $m_A(A) = \mathrm{O}$. Se tiene que $r(A) = \mathrm{O}$. Luego

$r(x)$ es el polinomio mínimo de A, lo cual contradice la hipótesis. Como $grad\big[r(x)\big] < grad\big[m_A(x)\big]$, se debe tener que $r(x) = 0$. Por lo tanto, $m_A(x)$ es un factor de $p_n(x)$. ■

Teorema 3.16 *Si A y B son matrices semejantes y $q_n(x)$ es cualquier polinomio, entonces:*

$$q_n\big(A\big) = Pq_n\big(B\big)P^{-1}.$$

Demostración.

Sean A y B matrices semejantes, luego existe una matriz P no singular tal que $A = PBP^{-1}$. Sea el polinomio de grado n

$$q_n(x) = c_n x^n + c_{n-1}x^{n-1} + \ldots + c_k x^k + \ldots + c_1 x + c_0.$$

Entonces,

$$\begin{aligned} q_n(A) &= c_n A^n + c_{n-1}A^{n-1} + \ldots + c_k A^k + \ldots + c_1 A + c_0 I \\ &= c_n\big(PBP^{-1}\big)^n + c_{n-1}\big(PBP^{-1}\big)^{n-1} + \ldots + c_k\big(PBP^{-1}\big)^k + \ldots \\ &\quad + c_1\big(PBP^{-1}\big) + c_0 I \\ &= c_n\big(PB^nP^{-1}\big) + c_{n-1}\big(PB^{n-1}P^{-1}\big) + \ldots + c_k\big(PB^kP^{-1}\big) + \ldots \\ &\quad + c_1\big(PBP^{-1}\big) + c_0\big(PP^{-1}\big) \\ &= P\big(c_nB^n + c_{n-1}B^{n-1} + \ldots + c_kB^k + \ldots + c_1B + c_0I\big)P^{-1} \\ &= P\big[q_n\big(B\big)\big]P^{-1}, \end{aligned}$$

como se quería demostrar. ■

Teorema 3.17 *Matrices semejantes tienen el mismo polinomio mínimo.*

Demostración.

Sea $m_A(x)$ el polinomio mínimo de A. Por el Teorema 3.16, se tiene que

$$m_A(A) = P\big[m_A\big(B\big)\big]P^{-1} = \mathrm{O}\,.$$

Esto implica que $m_A(B) = \mathrm{O}$, es decir $m_A(x)$ es también el polinomio mínimo de B, porque por el Teorema 3.15 si existiera un divisor que se

anulase para B, también se anularía para A, lo que contradice la hipótesis de que $m_A(x)$ es el polinomio mínimo de A. Luego los polinomios mínimos de A y B son iguales. ■

El siguiente teorema establece una importante relación entre el polinomio característico y el polinomio mínimo.

Teorema 3.18 *Si A satisface un polinomio $p_n(x)$, entonces todo valor propio de A es también raíz de $p_n(x)$. Por consiguiente, todo valor propio de A es una raíz del polinomio mínimo de A.*

Demostración.

Supóngase que A satisface $p_n(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 x^0$. Si λ es un valor propio de A, entonces se puede encontrar un vector propio $\vec{v}$ tal que $A\vec{v} = \lambda\vec{v}$. Así, $A^2\vec{v} = A\big(A\vec{v}\big) = A\big(\lambda\vec{v}\big) = \lambda\big(A\vec{v}\big) = \lambda\big(\lambda\vec{v}\big) = \lambda^2\vec{v}$. Si se continúa de esta manera, se origina

$$A^k\vec{v} = \lambda^k\vec{v}, \qquad \text{para todo } k > 0.$$

Pero como $p_n(A) = 0$, se tiene

$$\begin{aligned}
\vec{0} &= p_n(A)\vec{v} = \big(a_n A^n + a_{n-1}A^{n-1} + \ldots + a_1 A + a_0 I\big)\vec{v} \\
&= \big(a_n A^n\big)\vec{v} + \big(a_{n-1}A^{n-1}\big)\vec{v} + \ldots + \big(a_1 A\big)\vec{v} + \big(a_0 I\big)\vec{v} \\
&= a_n\lambda^n\vec{v} + a_{n-1}\lambda^{n-1}\vec{v} + \ldots + a_1\lambda\vec{v} + a_0\vec{v} \\
&= \big(a_n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0\lambda^0\big)\vec{v} = p_n\big(\lambda\big)\vec{v},
\end{aligned}$$

como $p_n(\lambda)$ es un escalar y $p_n(\lambda)\vec{v} = \vec{0}$ con $\vec{v} \neq \vec{0}$, se concluye que $p_n(\lambda) = 0$. Por ello, λ es una raíz de $p_n(x)$.

Puesto que A satisface su polinomio mínimo, todo valor propio de A es raíz de dicho polinomio mínimo. ■

Ejemplo 3.8 *Encuentre el polinomio minimal que satisface la matriz dada en el Ejemplo 2.12.*

Solución.

En el Ejemplo 2.12, se obtuvo que los valores propios de A eran $\lambda_1 = 0$ y $\lambda_2 = 2$ (de multiplicidad algebraica 2). Como la multiplicidad algebraica

del valor propio λ_2 resultó ser igual a su multiplicidad geométrica, el polinomio

$$p(x) = \big(x - 0\big)\big(x - 2\big) = x^2 - 2x$$

es satisfecho por la matriz A. Veamos que

$$p\big(A\big) = A^2 - 2A = \mathrm{O},$$

al sustituir la matriz A se tiene que

$$\begin{aligned} A^2 - 2A &= \begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix}^2 - 2\begin{bmatrix} 3 & 2 & -3 \\ -3 & -4 & 9 \\ -1 & -2 & 5 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 4 & -6 \\ -6 & -8 & 18 \\ -2 & -4 & 10 \end{bmatrix} - \begin{bmatrix} 6 & 4 & -6 \\ -6 & -8 & 18 \\ -2 & -4 & 10 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Ejemplo 3.9 *Encuentre el polinomio minimal que satisface la matriz dada en el Ejemplo 2.18.*

Solución.

En el Ejemplo 2.18, se obtuvo que los valores propios de A eran $\lambda_1 = 1$ (de multiplicidad algebraica 2) y $\lambda_2 = 7$. Como se obtuvo que la multiplicidad algebraica del valor propio λ_1 era igual a su multiplicidad geométrica, el polinomio minimal es

$$m_A(x) = \big(x - 1\big)\big(x - 7\big) = x^2 - 8x + 7.$$

El lector puede verificar que la matriz A satisface este polinomio.

3.4 Forma canónica de Jordan

Ahora se considera una versión completa del teorema de Cayley−Hamilton, el cual fue estudiado en la sección anterior. Esta nueva versión será usada para la forma canónica de Jordan que estudiaremos ahora. La forma de Jordan utiliza todo el material estudiado en los capítulos precedentes. El lector que desee profundizar en este tema puede consultar Weintraub (2009).

Definición 3.6 Bloque de Jordan

Una matriz triangular superior de tamaño $r \times r$, $\mathcal{J}_r(\lambda)$, es un bloque elemental de Jordan si se verifica que:

i) Todos sus elementos en la diagonal principal son iguales a λ.

ii) Todos sus elementos en la primera sobrediagonal son iguales a 1.

iii) Todos los demás elementos son iguales a 0.

De este modo, $\mathcal{J}_r(\lambda)$ es de la forma:

$$\mathcal{J}_r(\lambda) = \begin{bmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda & 1 \\ 0 & 0 & \ldots & 0 & \lambda \end{bmatrix} = \lambda I_r + N_r \quad con \quad N_r = \begin{bmatrix} 0 & 1 & \ldots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \ldots & 0 \end{bmatrix},$$

donde N_r es una matriz nilpotente de índice r.

Nótese que las matrices λI_r y N_r conmutan entre sí. Como una matriz de Jordan está constituida por bloques elementales, su definición es la que sigue.

Definición 3.7 Matriz de Jordan

Una matriz $\mathcal{J}$ de tamaño $n \times n$, de la forma

$$\mathcal{J} = \begin{bmatrix} \mathcal{J}_{n_1}(\lambda_1) & 0 & 0 & \cdots & 0 \\ 0 & \mathcal{J}_{n_2}(\lambda_2) & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & \mathcal{J}_{n_{s-1}}(\lambda_{s-1}) & 0 \\ 0 & \cdots & \cdots & 0 & \mathcal{J}_{n_s}(\lambda_s) \end{bmatrix},$$

en donde $\mathcal{J}_{n_1}, \mathcal{J}_{n_2}, \ldots, \mathcal{J}_{n_s}$ *son bloques elementales de Jordan de órdenes* $n_1 \geq n_2 \geq \cdots \geq n_s$ *con* $s \geq 1$*, se denomina matriz de Jordan.*

Teorema 3.19 *Sea* A *una matriz real de tamaño* $n \times n$*. Entonces existe una matriz* P *no singular tal que*

$$P^{-1}AP = \begin{bmatrix} \mathcal{J}_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \mathcal{J}_{n_\mu} \end{bmatrix} = \mathcal{J}, \tag{3.11}$$

en donde cada $\mathcal{J}_{n_k}$ *es un bloque de Jordan de tamaño* $n_k \times n_k$ *y el subíndice* $\mu = \mu_1 + \mu_2 + \ldots + \mu_s$ *es igual a la suma de las multiplicidades geométricas de los valores propios distintos de* A*. Un mismo valor propio* λ_k *puede estar en distintos bloques de Jordan* $\mathcal{J}_{n_k}$*, pero el número total de bloques con ese valor propio es igual a su multiplicidad geométrica* μ_k*, mientras que el número total de elementos en la diagonal principal con ese valor propio es igual a su multiplicidad algebraica* m_k*. Los números* n_k *y el número total de bloques quedan determinados de manera única por la matriz* A*.*

Demostración.

Para identificar los pasos a seguir en la demostración, supongamos que A es una matriz real de tamaño 2×2 que tiene un único valor propio λ. Sea $\vec{u}_1$ el único vector propio correspondiente a λ. Entonces A no es diagonalizable. Veamos que

$$P^{-1}AP = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}.$$

Como la multiplicidad algebraica es diferente de la multiplicidad geométrica, se encuentra un vector propio generalizado $\vec{u}_2$.

Por el Teorema 2.33, los vectores $\vec{u}_1$ y $\vec{u}_2$ son linealmente independientes, luego $P = \left[\begin{array}{cc} \vec{u}_1 & \vec{u}_2 \end{array}\right]$ es no singular. Por lo tanto,

$$AP = A\left[\begin{array}{cc} \vec{u}_1 & \vec{u}_2 \end{array}\right] = \left[\begin{array}{cc} A\vec{u}_1 & A\vec{u}_2 \end{array}\right] = \left[\begin{array}{cc} \lambda\vec{u}_1 & A\vec{u}_2 \end{array}\right].$$

Pero de la ecuación (2.37), se tiene que $A\vec{u}_2 = \lambda\vec{u}_2 + \vec{u}_1$, de manera que

$$AP = \left[\begin{array}{cc} \lambda\vec{u}_1 & \lambda\vec{u}_2 + \vec{u}_1 \end{array}\right].$$

Por otra parte,

$$\begin{aligned} P\mathcal{J} &= \left[\begin{array}{cc} \vec{u}_1 & \vec{u}_2 \end{array}\right] \left[\begin{array}{cc} \lambda & 1 \\ 0 & \lambda \end{array}\right] \\ &= \left[\begin{array}{cc} \lambda\vec{u}_1 & \vec{u}_1 + \lambda\vec{u}_2 \end{array}\right]. \end{aligned}$$

Por lo tanto, $AP = P\mathcal{J}$, lo que significa que $P^{-1}AP = \mathcal{J}$. Luego, el teorema es válido para matrices de tamaño 2×2.

Para probar el teorema para el caso general, se escribe P en términos de sus columnas como

$$P = \left[\begin{array}{cccc} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{array}\right],$$

donde algunos $\vec{u}_j$ son vectores propios generalizados. Consideremos $\vec{u}_i$ como el elemento de una *hilera de vectores* encabezados por algún vector propio $\vec{u}_{i-1}$ y determinados por

$$A\vec{u}_i = \lambda_i\vec{u}_i + \theta_i\vec{u}_{i-1}, \tag{3.12}$$

donde $\vec{u}_0 = \vec{0}$, λ_i es el valor propio en el bloque de Jordan que afecta a $\vec{u}_i$ y

$$\theta_i = \begin{cases} 0 & \text{si } \vec{u}_i \text{ es un vector propio,} \\ 1 & \text{si } \vec{u}_i \text{ es un vector propio generalizado.} \end{cases}$$

Nótese que si $\theta_i = 1$, entonces $\vec{u}_i$ es el único vector adicional en la hilera y el bloque correspondiente de Jordan es de tamaño 2×2. Así, cada hilera produce un solo bloque en la matriz $\mathcal{J}$. Luego, la clave para encontrar la forma de Jordan de A se convierte en la búsqueda de las hileras de los vectores definidos en (3.12).

Básicamente, se tiene que mostrar de qué manera se pueden construir estas hileras para cada matriz $A \in \mathcal{M}_{nn}$.

Empleando la metodología dada en Filippov (1971), se procede por inducción matemática, partiendo del hecho de que cada matriz de tamaño 1×1 está ya en su forma de Jordan. La prueba consiste en suponer que

se logra la construcción para todas las matrices de orden menor que n (esta es la "hipótesis de inducción"), y después se aplican tres pasos para obtener la forma de Jordan de una matriz de tamaño $n \times n$. Los pasos que se aplican son:

(i) Se supone que A es singular, entonces su espacio columna $\big[Col(A)\big]$ tiene dimensión $p < n$. En lo que respecta solamente a este espacio pequeño, la hipótesis de inducción garantiza que una forma de Jordan es posible; luego, deben existir p vectores linealmente independientes $\vec{u}_i \in Col(A)$ tales que

$$A\vec{u}_i = \lambda_i \vec{u}_i + \theta_i \vec{u}_{i-1}. \tag{3.13}$$

(ii) Se asume que el espacio nulo $\big[\ker(A)\big]$ y $Col(A)$ tienen una intersección de dimensión q. Luego, cada vector del $\ker(A)$ es también un vector propio correspondiente al valor propio $\lambda = 0$. Por lo tanto, se tienen q hileras en el paso (i) que comienzan a partir de este valor propio, y nos interesan los vectores $\vec{u}_i$ que están al final de dichas hileras. Puesto que cada uno de estos q vectores está en $Col(A)$, estos se pueden expresar como una combinación de las columnas de A, es decir:

$$\vec{u}_i = A\vec{v}_i, \qquad \text{para algún} \qquad \vec{v}_i \in \mathbb{R}^n.$$

(iii) Se considera que $\nu\big(A\big) = \dim\big[\ker(A)\big] = n - p$. Entonces, independientemente de su intersección q-dimensional con $Col(A)$, debe contener $n - p - q$ vectores básicos adicionales $\vec{w}_i$ fuera de esa intersección.

Al juntar estos pasos, se obtiene el teorema de Jordan.

Los p vectores $\vec{u}_i$, los q vectores $\vec{v}_i$ y los $n - p - q$ vectores $\vec{w}_i$ forman las hileras de Jordan para la matriz A y estos vectores son linealmente independientes. Si estos vectores conforman las columnas de la matriz P, entonces P es no singular y $P^{-1}AP = \mathcal{J}$ está en la forma de Jordan.

Si se quieren renumerar estos vectores como $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ y hacerlos corresponder con las condiciones de la ecuación (3.12), entonces cada $\vec{v}_i$ se debe colocar inmediatamente después del $\vec{u}_i$ del cual se origina; esto completa una hilera en la cual $\lambda_i = 0$. Los $\vec{w}_i$ vienen al final, cada uno solo en su propia hilera; de nuevo el valor propio es cero, ya que los $\vec{w}_i$ están en el $\ker(A)$. Los bloques que tienen valores propios diferentes de cero se completaron en el paso (i), los bloques con valores propios iguales

a cero crecen en una fila y una columna en el paso (ii), y el paso (iii) contribuye con cualquiera de los bloques de tamaño 1×1, $\mathcal{J}_i = [0]$.

En esta construcción, el único punto técnico es verificar la independencia de toda la colección $\vec{u}_i$, $\vec{v}_i$ y $\vec{w}_i$. Supongamos que alguna combinación de estos es cero, es decir

$$\sum_{i=1}^{p} \alpha_i \vec{u}_i + \sum_{i=1}^{q} \beta_i \vec{v}_i + \sum_{i=1}^{n-p-q} \gamma_i \vec{w}_i = \vec{0}. \tag{3.14}$$

Si se premultiplica por A y se usan las relaciones dadas en (3.13) para $\vec{u}_i$

$$\sum_{i=1}^{p} \alpha_i \big[\lambda_i \vec{u}_i + \theta_i \vec{u}_{i-1}\big] + \sum_{i=1}^{q} \beta_i \underbrace{A\vec{w}_i}_{\vec{v}_i} = \vec{0}, \tag{3.15}$$

como los $A\vec{v}_i$ son los $\vec{u}_i$ especiales al final de las hileras correspondientes a $\lambda_i = 0$, no pueden aparecer en la primera sumatoria. Luego, (3.15) es una combinación de los $\vec{u}_i$, los cuales son independientes por la hipótesis de inducción (proporcionan la forma de Jordan en $Col(A)$). Por lo tanto, se concluye que cada β_i es cero. Si se reemplaza en (3.14), se llega a

$$\sum_{i=1}^{p} \alpha_i \vec{u}_i = - \sum_{i=1}^{n-p-q} \gamma_i \vec{w}_i.$$

Como el lado izquierdo está en $Col(A)$ y los $\vec{w}_i$ son independientes de ese espacio, cada γ_i debe ser cero. Por lo tanto,

$$\sum_{i=1}^{p} \alpha_i \vec{u}_i = \vec{0}$$

y de la independencia de los $\vec{u}_i$, se tiene que todos los $\alpha_i = 0$.

Si la matriz A inicial no es singular, se pueden aplicar los tres pasos a $\widetilde{A} = A - \kappa I_n$ (si se elige la constante κ de manera que $\widetilde{A}$ sea singular y que pueda ser cualquiera de los valores propios de A). El algoritmo pone $\widetilde{A}$ en su forma de Jordan $P^{-1}\widetilde{A}P = \widetilde{\mathcal{J}}$ al generar las hileras $\vec{u}_i$ de las $\vec{u}_i$, $\vec{v}_i$ y $\vec{w}_i$. Entonces, la forma de Jordan de A utiliza las mismas hileras y la misma P

$$P^{-1}AP = P^{-1}\big(\widetilde{A} + \kappa I_n\big)P = \widetilde{\mathcal{J}} + P^{-1}\kappa P = \widetilde{\mathcal{J}} + \kappa I_n = \mathcal{J}.$$

Esto completa la prueba de que cada A es semejante a alguna matriz de Jordan $\mathcal{J}$. Excepto por el reordenamiento de los bloques, *es semejante a solo una* $\mathcal{J}$. En este sentido, A tiene una forma de Jordan única. ■

Definición 3.8 Forma canónica de Jordan

La matriz $\mathcal{J}$ dada en el Teorema 3.19 se denomina forma canónica de Jordan de A.

Ejemplo 3.10

Encuentre una matriz P no singular tal que la matriz A dada en el Ejemplo 2.22 sea semejante a una matriz de Jordan.

Solución.

En el Ejemplo 2.22, se obtuvo que el único valor propio de A era $\lambda = 2$ de multiplicidad algebraica 3 y además los siguientes vectores propios generalizados para la matriz A:

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}.$$

Por tanto,

$$P = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 3 \\ -1 & 0 & 0 \end{bmatrix} \qquad \text{y} \qquad P^{-1} = \begin{bmatrix} 0 & 0 & -1 \\ 3 & -2 & -2 \\ -1 & 1 & 1 \end{bmatrix}.$$

Al realizar el producto $P^{-1}AP$, se obtiene

$$\begin{bmatrix} 0 & 0 & -1 \\ 3 & -2 & -2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ -3 & 2 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 3 \\ -1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Nótese que en la diagonal de la matriz de Jordan, aparecen los valores propios de A.

Ejemplo 3.11

Encuentre una matriz P no singular tal que $P^{-1}AP$ sea una matriz de Jordan, para la matriz A dada en el Ejemplo 3.5.

Solución.

En el Ejemplo 3.5, se obtuvo que $\lambda = 1$ era el único valor propio con multiplicidad algebraica igual a tres, y además se encontraron los siguientes vectores propios: $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix}$. Para encontrar un vector propio generalizado $\vec{v}_3$, se calcula

$$\begin{aligned} (A - I)\vec{v}_3 &= c_1\vec{v}_1 + c_2\vec{v}_2 \\ \begin{bmatrix} 4 & -3 & -2 \\ 8 & -6 & -4 \\ -4 & 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} &= c_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix}. \end{aligned}$$

Al realizar las operaciones por filas, se obtiene

$$\left[\begin{array}{ccc|c} 4 & -3 & -2 & c_1 \\ 8 & -6 & -4 & 2c_2 \\ -4 & 3 & 2 & 2c_1 - 3c_2 \end{array}\right] \underset{\substack{F_2 - 2F_1 \\ F_3 + F_2}}{\sim} \left[\begin{array}{ccc|c} 4 & -3 & -2 & c_1 \\ 0 & 0 & 0 & 2c_2 - 2c_1 \\ 0 & 0 & 0 & 3c_1 - 3c_2 \end{array}\right].$$

Para que el sistema tenga infinitas soluciones, $c_1 = c_2$; luego, $4x - 3y - 2z = c_1$. Por lo tanto, $z = 2x - \frac{3}{2}y - \frac{1}{2}c_1$. Si se hace $x = y = 0$ y $c_1 = 1$, se obtiene el vector propio generalizado: $\vec{v}_3 = \frac{1}{2}\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$. Por consiguiente,

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 2 & -1 & -\frac{1}{2} \end{bmatrix} \qquad \text{y} \qquad P^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 4 & -3 & -2 \end{bmatrix}.$$

Obsérvese que la segunda columna de P es una combinación lineal de los dos vectores propios que constituyen la base del espacio propio asociado al valor propio $\lambda = 1$. Si se realiza el producto $P^{-1}AP$, se llega a

$$\begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 4 & -3 & -2 \end{bmatrix} \begin{bmatrix} 5 & -3 & -2 \\ 8 & -5 & -4 \\ -4 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 2 & -1 & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

El lector puede notar que sobre la diagonal de la matriz de Jordan se encuentra el respectivo valor propio de la matriz A.

Ejercicios 3.2

1. Sean $p_2(x) = x^2 - 2x - 3$, $q_2(x) = x^3 - 3x - 2$ y las matrices

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 1 & -1 \\ 3 & -1 \end{bmatrix}.$$

Determine:

a. $p_2(A)$. *b.* $p_2(B)$. *c.* $q_2(A)$. *d.* $q_2(B)$.

e. $p_2(A+B)$. *f.* $p_2(AB)$. *g.* $q_2(A+B)$. *h.* $q_2(AB)$.

2. Calcular el polinomio mínimo y la descomposición de Jordan de las siguientes matrices:

$$a. \begin{bmatrix} 6 & -1 \\ 1 & 4 \end{bmatrix}. \qquad b. \begin{bmatrix} -8 & -7 \\ 7 & 6 \end{bmatrix}. \qquad c. \begin{bmatrix} 4 & -1 \\ 1 & 2 \end{bmatrix}.$$

$$d. \begin{bmatrix} a & -b \\ b & c \end{bmatrix}. \qquad e. \begin{bmatrix} 3 & -1 & -1 \\ 6 & -2 & -3 \\ -2 & 1 & 2 \end{bmatrix}. \qquad f. \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ -2 & -3 & -2 \end{bmatrix}.$$

3. Encuentre todas las respectivas formas canónicas de Jordan para las matrices cuyos polinomios característicos $p_A(x)$ y minimales $m_A(x)$ son los siguientes:

a. $p_A(x) = (x-1)^4(x-3)^2$, $m_A(x) = (x-1)^2(x-3)^2$.

b. $p_A(x) = (x-5)^4$, $m_A(x) = (x-5)^2$.

c. $p_A(x) = (x-2)^4(x-4)^4$, $m_A(x) = (x-2)^2(x-4)^2$.

4. Demuestre que la m-ésima potencia de la matriz de Jordan es:

$$\mathcal{J}^m = \begin{bmatrix} \mathcal{J}_{n_1}^m(\lambda_1) & 0 & 0 & \cdots & 0 \\ 0 & \mathcal{J}_{n_2}^m(\lambda_2) & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & \mathcal{J}_{n_{s-1}}^m(\lambda_{s-1}) & 0 \\ 0 & \cdots & \cdots & 0 & \mathcal{J}_{n_s}^m(\lambda_s) \end{bmatrix},$$

donde la m-ésima potencia de un bloque de Jordan de tamaño $n_k \times n_k$, para $m \leq n_{k-1}$, es dada por

$$\mathcal{J}_{n_k}^m(\lambda_k) = \begin{bmatrix} \lambda_k^m & \binom{m}{1}\lambda_k^{m-1} & \binom{m}{2}\lambda_k^{m-2} & \cdots & \binom{m}{n_k-1}\lambda_k^{m-n_k+1} \\ 0 & \lambda_k^m & \binom{m}{1}\lambda_k^{m-1} & \cdots & \binom{m}{n_k-2}\lambda_k^{m-n_k+2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^m & \binom{m}{1}\lambda_k^{m-1} \\ 0 & 0 & \cdots & 0 & \lambda_k^m \end{bmatrix}.$$

3.5 Raíces cuadradas

Si se considera un número $a \in \mathbb{R}$, como $\big(-a\big)^2 = a^2$, es evidente que para cualquier $a > 0$, se tienen dos raíces cuadradas; una positiva y otra negativa; mientras que cuando $a < 0$, sus raíces cuadradas son dos imaginarios puros, una raíz es la conjugada de la otra. En general, si $a \in \mathbb{C}$, también a tiene dos raíces cuadradas distintas. En esta sección se extiende el concepto de raíz cuadrada para estudiar la raíz cuadrada de una matriz de tamaño $n \times n$, tema poco trabajado en la mayoría de textos de álgebra lineal.

Definición 3.9 *Sea A una matriz real de tamaño $n \times n$. Una matriz X de tamaño $n \times n$ se llama raíz cuadrada de A si cumple que*

$$X^2 = A. \tag{3.16}$$

La matriz X puede tener algunos elementos complejos.

Cuando la matriz es de tamaño 2×2, existe un método sencillo para determinar sus raíces cuadradas, el cual fue presentado en Cayley (1858, p. 25) y es dado en el teorema que sigue.

Teorema 3.20 Método de Cayley

Sea A una matriz real de tamaño 2×2, entonces la raíz cuadrada de A viene dada por

$$A^{\frac{1}{2}} = \frac{1}{\sqrt{\operatorname{tr}(A) + 2\delta}}\big[A + \delta I_2\big], \tag{3.17}$$

donde $\delta = \pm\sqrt{\det\big(A\big)}$. Cuando $\operatorname{tr}(A) + 2\delta \neq 0$ entonces A tiene 4 raíces.

Demostración.

Sea $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$. Luego si la matriz dada en (3.17) es la raíz cuadrada de A, entonces $A = \left(A^{\frac{1}{2}}\right)\left(A^{\frac{1}{2}}\right)$

$$\left(A^{\frac{1}{2}}\right)^2 = \left(\frac{1}{\sqrt{\operatorname{tr}(A) \pm 2\sqrt{\det A}}}\right)^2 \left[\begin{array}{cc} a \pm \sqrt{\det A} & b \\ c & d \pm \sqrt{\det A} \end{array}\right]^2$$

$$= \frac{1}{\operatorname{tr}(A) \pm 2\sqrt{\det A}} \left[\begin{array}{cc} a^2 + ad \pm 2a\sqrt{\det A} & ba + bd \pm 2b\sqrt{\det A} \\ ca + cd \pm 2c\sqrt{\det A} & da + d^2 \pm 2d\sqrt{\det A} \end{array}\right].$$

De este modo,

$$\left(A^{\frac{1}{2}}\right)^2 = \frac{\big(a + d\big) \pm 2\sqrt{\det A}}{\operatorname{tr}(A) \pm 2\sqrt{\det A}} \left[\begin{array}{cc} a & b \\ c & d \end{array}\right].$$

Pero como $\operatorname{tr}(A) = a + d$, se tiene que

$$\left(A^{\frac{1}{2}}\right)^2 = A. \ \blacksquare$$

Ejemplo 3.12 *Determine para cada una de las siguientes matrices, una raíz cuadrada:*

$$A = \left[\begin{array}{cc} 2 & 1 \\ 0 & 2 \end{array}\right] \qquad y \qquad B = \left[\begin{array}{cc} 4 & 2 \\ -2 & 4 \end{array}\right].$$

Solución.

- Para la matriz A, se tiene que

$$\operatorname{tr}(A) = 4 \qquad \text{y} \qquad \det A = 4.$$

Como $\operatorname{tr}(A) = 2\sqrt{\det A}$, entonces A posee 2 raíces cuadradas. Una raíz cuadrada de A sería

$$A^{\frac{1}{2}} = \frac{1}{2\sqrt{2}} \begin{bmatrix} 2+2 & 1 \\ 0 & 2+2 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \frac{1}{4}\sqrt{2} \\ 0 & \sqrt{2} \end{bmatrix},$$

y multiplicando por -1, se obtiene la otra raíz.

El lector puede verificar que $\left(A^{\frac{1}{2}}\right)^2 = A$.

- Para la matriz B, se tiene que

$$\operatorname{tr}(B) = 8 \qquad \text{y} \qquad \det B = 20.$$

Como $\operatorname{tr}(B) \neq \pm 2\sqrt{\det B}$, entonces B posee 4 raíces cuadradas. Una raíz cuadrada de B es

$$B^{\frac{1}{2}} = \frac{1}{2\sqrt{2+\sqrt{5}}} \begin{bmatrix} 4+2\sqrt{5} & 2 \\ -2 & 4+2\sqrt{5} \end{bmatrix} = \begin{bmatrix} \sqrt{\sqrt{5}+2} & \sqrt{\sqrt{5}-2} \\ -\sqrt{\sqrt{5}-2} & \sqrt{\sqrt{5}+2} \end{bmatrix}.$$

Nótese que $-B^{\frac{1}{2}}$ también es raíz y otra raíz cuadrada de B es

$$B^{\frac{1}{2}} = \frac{1}{2\sqrt{2-\sqrt{5}}} \begin{bmatrix} 4-2\sqrt{5} & 2 \\ -2 & 4-2\sqrt{5} \end{bmatrix} = i \begin{bmatrix} \sqrt{\sqrt{5}-2} & -\sqrt{\sqrt{5}+2} \\ \sqrt{\sqrt{5}+2} & \sqrt{\sqrt{5}-2} \end{bmatrix}.$$

En este caso, $-B^{\frac{1}{2}}$ también es raíz; el lector puede veficarlo.

Teorema 3.21 *Sea $D = [d_{ii}]$ una matriz real diagonal de tamaño $n \times n$, entonces una raíz cuadrada de D es:*

$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{d_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{d_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{d_{nn}} \end{bmatrix}. \tag{3.18}$$

Demostración.
Queda como ejercicio para el lector. ■

Nota 3.2 Como cada elemento d_{ii} tiene dos raíces cuadradas $\sqrt{d_{ii}}$ y $-\sqrt{d_{ii}}$, entonces en la matriz (3.18), se puede reemplazar por la otra raíz del elemento d_{ii} y se obtiene una nueva raíz cuadrada para D.

Teorema 3.22 *Sea A una matriz real de tamaño $n \times n$ diagonalizable y sea P una matriz no singular tal que la matriz $D = P^{-1}AP$ es diagonal. Entonces una raíz cuadrada de A es*

$$A^{\frac{1}{2}} = PD^{\frac{1}{2}}P^{-1}, \tag{3.19}$$

donde $D^{\frac{1}{2}}$ es definida como en (3.18).

Demostración.
La demostración consiste en un cálculo directo,

$$\left(A^{\frac{1}{2}}\right)^2 = \left(PD^{\frac{1}{2}}P^{-1}\right)\left(PD^{\frac{1}{2}}P^{-1}\right) = P\left(D^{\frac{1}{2}}\right)^2P^{-1} = PDP^{-1} = A.$$

Así, queda el teorema probado. ■

Cuando son iguales la multiplicidad algebraica y la multiplicidad geométrica de los valores propios de una matriz A, se tiene que A es semejante a una matriz diagonal D cuyos elementos son los valores propios de A. Por lo tanto, si A es diagonalizable como cada valor propio tiene dos raíces cuadradas, entonces el número de raíces cuadradas de la matriz A es igual a 2^n. Si todos los valores propios de A son nulos, entonces A no tiene raíz cuadrada.

Ejemplo 3.13 *Determine las raíces cuadradas de la siguiente matriz:*

$$A = \begin{bmatrix} 12 & 1 & -2 & -11 \\ 7 & 11 & -5 & 12 \\ -10 & -3 & 16 & -1 \\ -3 & 4 & 7 & 15 \end{bmatrix}.$$

Solución.

Para la matriz A, se tiene que el polinomio característico es

$$p_A(\lambda) = \lambda^4 - 54\lambda^3 + 969\lambda^2 - 6676\lambda + 14400.$$

Entonces, sus valores propios son $\lambda_1 = 4$, $\lambda_2 = 9$, $\lambda_3 = 16$ y $\lambda_4 = 25$. La matriz A se puede expresar como

$$A = \begin{bmatrix} \frac{2}{3} & -\frac{23}{4} & \frac{115}{144} & -\frac{115}{252} \\ -1 & \frac{23}{4} & -\frac{115}{48} & 0 \\ \frac{1}{3} & -\frac{23}{4} & \frac{115}{72} & \frac{115}{252} \\ \frac{1}{3} & 0 & -\frac{115}{144} & \frac{115}{252} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 25 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 & 1 \\ -\frac{4}{23} & -\frac{4}{23} & -\frac{4}{23} & 0 \\ -\frac{48}{115} & -\frac{48}{115} & 0 & -\frac{48}{115} \\ -\frac{84}{115} & 0 & \frac{84}{115} & \frac{84}{115} \end{bmatrix}.$$

Como A tiene 4 valores propios distintos no nulos, entonces posee $2^4 = 16$ raíces cuadradas, al tomar todas las raíces positivas de los valores propios

$$A^{\frac{1}{2}} = \begin{bmatrix} \frac{2}{3} & -\frac{23}{4} & \frac{115}{144} & -\frac{115}{252} \\ -1 & \frac{23}{4} & -\frac{115}{48} & 0 \\ \frac{1}{3} & -\frac{23}{4} & \frac{115}{72} & \frac{115}{252} \\ \frac{1}{3} & 0 & -\frac{115}{144} & \frac{115}{252} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 & 1 \\ -\frac{4}{23} & -\frac{4}{23} & -\frac{4}{23} & 0 \\ -\frac{48}{115} & -\frac{48}{115} & 0 & -\frac{48}{115} \\ -\frac{84}{115} & 0 & \frac{84}{115} & \frac{84}{115} \end{bmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 10 & 1 & 0 & -5 \\ 3 & 9 & -3 & 6 \\ -4 & -1 & 12 & -1 \\ -1 & 2 & 3 & 11 \end{bmatrix}.$$

Para obtener las otras raíces cuadradas de la matriz A, se modifican los elementos de $D^{\frac{1}{2}}$ por las raíces negativas de los valores propios, como se muestra a continuación

$$A^{\frac{1}{2}} = \frac{1}{3} \begin{bmatrix} 10 & 9 & 8 & -13 \\ 3 & -3 & -15 & 18 \\ -4 & 3 & 16 & -5 \\ -1 & 6 & 7 & 7 \end{bmatrix} \quad \text{cuando} \quad D^{\frac{1}{2}} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix};$$

$$A^{\frac{1}{2}} = \frac{1}{3} \begin{bmatrix} -8 & -17 & -18 & -5 \\ 21 & 27 & 15 & 6 \\ -22 & -19 & -6 & -1 \\ -1 & 2 & 3 & 11 \end{bmatrix} \quad \text{tomando} \quad D^{\frac{1}{2}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix};$$

$$A^{\frac{1}{2}} = \begin{bmatrix} 6 & 3 & 0 & 1 \\ -7 & -5 & -1 & -6 \\ 4 & 5 & 4 & 5 \\ -3 & -2 & 1 & 1 \end{bmatrix} \quad \text{asumiendo} \quad D^{\frac{1}{2}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix};$$

$$A^{\frac{1}{2}} = \frac{1}{3}\begin{bmatrix} 0 & 1 & 10 & 5 \\ 3 & 9 & -3 & 6 \\ 6 & -1 & 2 & -11 \\ 9 & 2 & -7 & 1 \end{bmatrix} \quad \text{con} \quad D^{\frac{1}{2}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -5 \end{bmatrix}.$$

Nótese que en la matriz $D^{\frac{1}{2}}$, se ha modificado solo un valor propio; ahora, se consideran las raíces cuadradas de A cuando se cambian dos valores propios.

$$A^{\frac{1}{2}} = \frac{1}{3}\begin{bmatrix} -8 & -9 & -10 & -13 \\ 21 & 15 & 3 & 18 \\ -22 & -15 & -2 & -5 \\ -1 & 6 & 7 & 7 \end{bmatrix} \quad \text{cuando} \quad D^{\frac{1}{2}} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix};$$

$$A^{\frac{1}{2}} = \frac{1}{3}\begin{bmatrix} 18 & 17 & 8 & -5 \\ -21 & -27 & -15 & -6 \\ 12 & 19 & 16 & 11 \\ -9 & -2 & 7 & -1 \end{bmatrix} \quad \text{tomando} \quad D^{\frac{1}{2}} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix};$$

$$A^{\frac{1}{2}} = \begin{bmatrix} 0 & 3 & 6 & -1 \\ 1 & -1 & -5 & 6 \\ 2 & 1 & 2 & -5 \\ 3 & 2 & -1 & -1 \end{bmatrix} \quad \text{asumiendo} \quad D^{\frac{1}{2}} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -5 \end{bmatrix}.$$

Se puede verificar que estas 8 matrices y sus respectivas matrices negativas son las 16 raíces cuadradas de A.

Teorema 3.23 *Sea $T = [t_{ij}]$ una matriz triangular superior de tamaño $n \times n$, con a lo más un elemento nulo en la diagonal. Entonces, existe $T^{\frac{1}{2}} = [\tau_{ij}]$ y sus elementos cumplen que*

$$\tau_{ij} = \begin{cases} \sqrt{t_{ii}} & j = i, \\ \dfrac{t_{ij}}{\tau_{ii} + \tau_{jj}} & j = i+1, \\ \dfrac{1}{\tau_{ii} + \tau_{jj}} \left(t_{ij} - \displaystyle\sum_{k=i+1}^{j-1} \tau_{ik}\tau_{kj} \right) & j > i+1. \end{cases} \tag{3.20}$$

Demostración.

La demostración es por inducción sobre n. Vamos a probar que si los elementos dados en (3.20) son las entradas de la raíz cuadrada de T, entonces

$$T = \left(T^{\frac{1}{2}}\right)\left(T^{\frac{1}{2}}\right).$$

Si $n = 1$, $T = [t]$ es una matriz real de tamaño 1×1 que es triangular, luego

$$\left(T^{\frac{1}{2}}\right)^2 = \left[\ \sqrt{t}\ \right]^2 = [t] = T.$$

Para $n = 2$, se tiene que

$$\left(T^{\frac{1}{2}}\right)^2 = \begin{bmatrix} \sqrt{t_{11}} & \dfrac{t_{12}}{\sqrt{t_{11}} + \sqrt{t_{22}}} \\ 0 & \sqrt{t_{22}} \end{bmatrix}^2 = \begin{bmatrix} t_{11} & t_{12}\dfrac{\sqrt{t_{11}} + \sqrt{t_{22}}}{\sqrt{t_{11}} + \sqrt{t_{22}}} \\ 0 & t_{22} \end{bmatrix} = T;$$

supóngase que es cierto para todas las matrices triangulares de orden $n-1$, es decir que existe una matriz de tamaño $\left(n-1\right) \times \left(n-1\right)$,

$$T_{n-1}^{\frac{1}{2}} = \begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} & \dots & \tau_{1\,n-1} \\ 0 & \tau_{22} & \tau_{23} & \dots & \tau_{2\,n-1} \\ 0 & 0 & \tau_{33} & \dots & \tau_{3\,n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \tau_{n-1\ n-1} \end{bmatrix},$$

tal que $\left(T_{n-1}^{\frac{1}{2}}\right)^2 = T_{n-1}$. Como una matriz triangular T de orden n, se puede particionar como

$$T = \begin{bmatrix} T_{n-1} & \vdots & \vec{U} \\ \dots & . & \dots \\ \vec{0}^t & \vdots & t_{nn} \end{bmatrix}, \qquad \text{donde} \qquad \vec{U} = \begin{bmatrix} t_{1n} \\ t_{2n} \\ \vdots \\ t_{n-1\,n} \end{bmatrix},$$

por la hipótesis de inducción

$$T^{\frac{1}{2}} = \begin{bmatrix} T_{n-1}^{\frac{1}{2}} & \vdots & \vec{c} \\ \dots & . & \dots \\ \vec{0}^t & \vdots & \tau_{nn} \end{bmatrix} \qquad \text{con} \qquad \tau_{nn} = \sqrt{t_{nn}},$$

y con

$$\vec{c} = \begin{bmatrix} \tau_{1n} \\ \tau_{2n} \\ \vdots \\ \tau_{n-1n} \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau_{11}+\tau_{nn}} \left(t_{1n} - \sum_{k=2}^{n-1} \tau_{1k}\tau_{kn} \right) \\ \frac{1}{\tau_{22}+\tau_{nn}} \left(t_{2n} - \sum_{k=3}^{n-1} \tau_{2k}\tau_{kn} \right) \\ \vdots \\ \frac{t_{n-1\ n}}{\tau_{n-1\ n-1}+\tau_{nn}} \end{bmatrix}.$$

Luego,

$$\left(T^{\frac{1}{2}}\right)^2 = \begin{bmatrix} T_{n-1}^{\frac{1}{2}} & \vdots & \vec{c} \\ \cdots & . & \cdots \\ \vec{0}^t & \vdots & \tau_{nn} \end{bmatrix}^2 = \begin{bmatrix} T_{n-1} & \vdots & \left(T_{n-1}^{\frac{1}{2}} + \tau_{nn} I_{n-1}\right)\vec{c} \\ \cdots & . & \cdots\cdots\cdots \\ \vec{0}^t & \vdots & \tau_{nn}^2 \end{bmatrix}.$$

Nótese que $\left(T_{n-1}^{\frac{1}{2}} + \tau_{nn} I_{n-1}\right)\vec{c} = \vec{U}$,

$$\begin{bmatrix} (\tau_{11}+\tau_{nn}) & \tau_{12} & \cdots & \tau_{1n-1} \\ 0 & (\tau_{22}+\tau_{nn}) & \cdots & \tau_{2n-1} \\ 0 & 0 & \cdots & \tau_{3n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\tau_{n-1\ n-1}+\tau_{nn}) \end{bmatrix} \begin{bmatrix} \tau_{1n} \\ \tau_{2n} \\ \tau_{3n} \\ \vdots \\ \tau_{n-1n} \end{bmatrix} = \begin{bmatrix} t_{1n} \\ t_{2n} \\ t_{3n} \\ \vdots \\ t_{n-1n} \end{bmatrix}.$$

Esto completa la prueba. ■

Ejemplo 3.14 *Determine una raíz cuadrada para la siguiente matriz:*

$$A = \begin{bmatrix} 4 & 15 & 27 & 19 \\ 0 & 9 & 21 & 22 \\ 0 & 0 & 16 & 18 \\ 0 & 0 & 0 & 25 \end{bmatrix}.$$

Solución.

Usando el procedimiento descrito en (3.20), se tiene que

$$\tau_{11} = \sqrt{t_{11}} = 2, \qquad \tau_{22} = \sqrt{t_{22}} = 3, \qquad \tau_{33} = \sqrt{t_{33}} = 4,$$

$$\tau_{44} = \sqrt{t_{44}} = 5, \qquad \tau_{12} = \frac{t_{12}}{\tau_{11}+\tau_{22}} = 3, \qquad \tau_{23} = \frac{t_{23}}{\tau_{22}+\tau_{33}} = 3,$$

$$\tau_{34} = \frac{t_{34}}{\tau_{33}+\tau_{44}} = 2, \quad \tau_{13} = \frac{t_{13}-\big(\tau_{12}\tau_{23}\big)}{\tau_{11}+\tau_{33}} = 3, \quad \tau_{24} = \frac{t_{24}-\big(\tau_{23}\tau_{34}\big)}{\tau_{22}+\tau_{44}} = 2,$$

finalmente

$$\tau_{14} = \frac{t_{14}-\big(\tau_{12}\tau_{24}+\tau_{13}\tau_{34}\big)}{\tau_{11}+\tau_{44}} = 1,$$

luego,

$$A^{\frac{1}{2}} = \begin{bmatrix} 2 & 3 & 3 & 1 \\ 0 & 3 & 3 & 2 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 0 & 5 \end{bmatrix}.$$

El lector puede verificar que $\left(A^{\frac{1}{2}}\right)^2 = A$.

Teorema 3.24 *Sea A una matriz real de tamaño $n \times n$ con valores propios reales (a lo más uno igual a cero) y sea P una matriz no singular tal que la matriz, $T = P^{-1}AP$ es triangular. Entonces, una raíz cuadrada de A es,*

$$A^{\frac{1}{2}} = PT^{\frac{1}{2}}P^{-1}, \tag{3.21}$$

donde los elementos de $T^{\frac{1}{2}}$ están dados en (3.20).

Demostración.

La demostración consiste en un cálculo directo

$$\left(A^{\frac{1}{2}}\right)^2 = \big(PT^{\frac{1}{2}}P^{-1}\big)\big(PT^{\frac{1}{2}}P^{-1}\big) = P\big(T^{\frac{1}{2}}\big)^2P^{-1} = PTP^{-1} = A.$$

Así, queda el teorema probado. ■

Hasta este momento hemos considerado las raíces cuadradas de matrices diagonalizables o triangularizables, pero como todas las matrices no se pueden factorizar de esta manera, a continuación se presentan algunos métodos para obtener las raíces cuadradas de una matriz.

Teorema 3.25 *Si A es una matriz real de tamaño 3×3 con a lo más un valor propio nulo, entonces sus raíces cuadradas son*

$$A^{\frac{1}{2}} = \big[A + \alpha I\big]^{-1}\big[\beta A + \big(\sqrt{\lambda_1}\sqrt{\lambda_2}\sqrt{\lambda_3}\big)I\big], \tag{3.22}$$

donde $\alpha = \sum_{i=1}^{2}\sum_{j=i+1}^{3}\big(\sqrt{\lambda_i}\sqrt{\lambda_j}\big)$ $\;y\;$ $\beta = \sum_{k=1}^{3}\sqrt{\lambda_k}$.

El teorema es válido para cualquier matriz; sin embargo, la prueba que se presenta en este material es únicamente para matrices triangularizables.

Demostración.

Supongamos que A es una matriz cuadrada con valores propios reales. Por el Teorema 3.7, es semejante a una matriz triangular superior T, luego puede expresarse como

$$A = QTQ^t, \tag{3.23}$$

donde Q es una matriz ortogonal y

$$T = \begin{bmatrix} \lambda_1 & a & b \\ 0 & \lambda_2 & c \\ 0 & 0 & \lambda_3 \end{bmatrix}. \tag{3.24}$$

Con $\lambda_1, \lambda_2, \lambda_3$ los valores propios de A. Al reemplazar (3.23) en (3.22), se tiene que

$$\begin{aligned} A^{\frac{1}{2}} &= \big[QTQ^t + \alpha I\big]^{-1}\big[\beta\big(QTQ^t\big) + \sqrt{\det\big(QTQ^t\big)}I\big] \\ &= \big[Q\big(T + \alpha I\big)Q^t\big]^{-1}\big[Q\big(\beta T + \sqrt{\det T}I\big)Q^t\big] \\ &= Q\big[T + \alpha I\big]^{-1}Q^tQ\big[\beta T + \sqrt{\det T}I\big]Q^t \\ &= Q\big[T + \alpha I\big]^{-1}\big[\beta T + \sqrt{\det T}I\big]Q^t. \end{aligned}$$

Es decir, al utilizar la descomposición de Schur, se llega a que

$$A^{\frac{1}{2}} = QT^{\frac{1}{2}}Q^t.$$

Luego, se debe demostrar que

$$T^{\frac{1}{2}} = \big[T + \alpha I\big]^{-1}\big[\beta T + \sqrt{\det T}I\big], \tag{3.25}$$

Para ello, se define

$$\lambda_{ij} = \sqrt{\lambda_i} + \sqrt{\lambda_j}. \tag{3.26}$$

Esto permite factorizar α y β de la siguiente manera

$$\begin{aligned}
\alpha &= \sqrt{\lambda_1}\sqrt{\lambda_2} + \sqrt{\lambda_1}\sqrt{\lambda_3} + \sqrt{\lambda_2}\sqrt{\lambda_3} = \lambda_{12}\,\lambda_{13} - \lambda_1 \\
&= \lambda_{12}\,\lambda_{23} - \lambda_2 = \lambda_{13}\,\lambda_{23} - \lambda_3 \\
\beta &= \lambda_{12} + \sqrt{\lambda_3} = \frac{\lambda_{12}\,\lambda_{13} - \sqrt{\lambda_2\lambda_3}}{\sqrt{\lambda_1}} \\
&= \frac{\lambda_{12}\,\lambda_{23} - \sqrt{\lambda_1\lambda_3}}{\sqrt{\lambda_2}} = \frac{\lambda_{13}\,\lambda_{23} - \sqrt{\lambda_1\lambda_2}}{\sqrt{\lambda_3}}.
\end{aligned}$$

Por lo tanto,

$$\begin{aligned}
T + \alpha I &= \begin{bmatrix} \lambda_1 & a & b \\ 0 & \lambda_2 & c \\ 0 & 0 & \lambda_3 \end{bmatrix} + \left(\sqrt{\lambda_1\lambda_2} + \sqrt{\lambda_1\lambda_3} + \sqrt{\lambda_2\lambda_3}\right) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \lambda_{12}\,\lambda_{13} & a & b \\ 0 & \lambda_{12}\,\lambda_{23} & c \\ 0 & 0 & \lambda_{13}\,\lambda_{23} \end{bmatrix},
\end{aligned}$$

y si se hace $\xi = \lambda_{12}\,\lambda_{13}\,\lambda_{23}$, entonces

$$T + \alpha I = \xi \begin{bmatrix} \frac{1}{\lambda_{23}} & \frac{a}{\xi} & \frac{b}{\xi} \\ 0 & \frac{1}{\lambda_{13}} & \frac{c}{\xi} \\ 0 & 0 & \frac{1}{\lambda_{12}} \end{bmatrix},$$

luego su inversa es

$$\left(T + \alpha I\right)^{-1} = \frac{1}{\xi} \begin{bmatrix} \lambda_{23} & -\frac{a}{\lambda_{12}} & \frac{ac}{\xi} - \frac{b}{\lambda_{13}} \\ 0 & \lambda_{13} & -\frac{c}{\lambda_{23}} \\ 0 & 0 & \lambda_{12} \end{bmatrix}. \tag{3.27}$$

Por otra parte,

$$\beta T + \sqrt{\det T}\, I = \left(\sqrt{\lambda_1} + \sqrt{\lambda_2} + \sqrt{\lambda_3}\,\right) \begin{bmatrix} \lambda_1 & a & b \\ 0 & \lambda_2 & c \\ 0 & 0 & \lambda_3 \end{bmatrix} + \sqrt{\lambda_1\lambda_2\lambda_3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

la cual en términos de ξ, se puede expresar como

$$\beta T + \sqrt{\det T}\, I = \xi \begin{bmatrix} \frac{\sqrt{\lambda_1}}{\lambda_{23}} & a\frac{\beta}{\xi} & b\frac{\beta}{\xi} \\ 0 & \frac{\sqrt{\lambda_2}}{\lambda_{13}} & c\frac{\beta}{\xi} \\ 0 & 0 & \frac{\sqrt{\lambda_3}}{\lambda_{12}} \end{bmatrix}. \tag{3.28}$$

Al reemplazar las matrices obtenidas en (3.27) y (3.28), en la expresión (3.25) se tiene

$$T^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & \frac{a}{\lambda_{12}} & \frac{b}{\lambda_{13}} - \frac{ac}{\xi} \\ 0 & \sqrt{\lambda_2} & \frac{c}{\lambda_{23}} \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix}.$$

El lector puede verificar que los elementos de esta última matriz coinciden con los dados en el Teorema 3.23. ■

Ejemplo 3.15 *Determine las raíces cuadradas para la matriz dada a continuación:*

$$A = \begin{bmatrix} 3 & -1 & 2 \\ -1 & 2 & -1 \\ -1 & 1 & 0 \end{bmatrix}.$$

Solución.

El polinomio característico de la matriz A es

$$p_A(\lambda) = -\lambda^3 + 5\lambda^2 - 8\lambda + 4.$$

De esta suma, se tiene que

$$\lambda_1 = 1, \qquad \lambda_2 = 2 \qquad \text{y} \qquad \lambda_3 = 2.$$

Como A tiene 2 valores propios distintos no nulos, entonces posee $2^2 = 4$ raíces cuadradas. Si se consideran las raíces positivas de los valores propios, se tiene que

$$\alpha = 2 + 2\sqrt{2} \qquad \text{y} \qquad \beta = 1 + 2\sqrt{2}.$$

Por lo tanto, una raíz cuadrada de A es

$$A^{\frac{1}{2}} = \big[A + \big(2 + 2\sqrt{2}\big)I\big]^{-1}\big[\big(1 + 2\sqrt{2}\big)A + \sqrt{4}I\big],$$

al sustituir se tiene que

$$A^{\frac{1}{2}} = \big(1+2\sqrt{2}\big)\begin{bmatrix} 5+2\sqrt{2} & -1 & 2 \\ -1 & 4+2\sqrt{2} & -1 \\ -1 & 1 & 2+2\sqrt{2} \end{bmatrix}^{-1}\begin{bmatrix} \frac{19+4\sqrt{2}}{7} & -1 & 2 \\ -1 & \frac{12+4\sqrt{2}}{7} & -1 \\ -1 & 1 & \frac{4\sqrt{2}-2}{7} \end{bmatrix}$$

$$= \frac{4\sqrt{2}-5}{8}\begin{bmatrix} 3+2\sqrt{2} & 4-2\sqrt{2} & 2\sqrt{2}-5 \\ 1 & 4+2\sqrt{2} & 1 \\ 1 & 2\sqrt{2}-4 & 9 \end{bmatrix}\begin{bmatrix} \frac{19+4\sqrt{2}}{7} & -1 & 2 \\ -1 & \frac{12+4\sqrt{2}}{7} & -1 \\ -1 & 1 & \frac{4\sqrt{2}-2}{7} \end{bmatrix}$$

$$= \frac{\sqrt{2}}{4}\begin{bmatrix} 5 & 2\sqrt{2}-4 & 5-2\sqrt{2} \\ -1 & 4 & -1 \\ -1 & 4-2\sqrt{2} & 2\sqrt{2}-1 \end{bmatrix}.$$

Nótese que $-A^{\frac{1}{2}}$ también es raíz, verifiquemos que en efecto son raíces cuadradas

$$\left(\pm A^{\frac{1}{2}}\right)^2 = \frac{2}{16}\begin{bmatrix} 5 & 2\sqrt{2}-4 & 5-2\sqrt{2} \\ -1 & 4 & -1 \\ -1 & 4-2\sqrt{2} & 2\sqrt{2}-1 \end{bmatrix}\begin{bmatrix} 5 & 2\sqrt{2}-4 & 5-2\sqrt{2} \\ -1 & 4 & -1 \\ -1 & 4-2\sqrt{2} & 2\sqrt{2}-1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -1 & 2 \\ -1 & 2 & -1 \\ -1 & 1 & 0 \end{bmatrix}.$$

Si se consideran las raíces negativas de los valores propios iguales, se tiene que

$$\alpha = 2-2\sqrt{2} \qquad \text{y} \qquad \beta = 1-2\sqrt{2}.$$

Por lo tanto, otra raíz cuadrada de A es

$$A^{\frac{1}{2}} = \big[A+\big(2-2\sqrt{2}\big)I\big]^{-1}\big[\big(1-2\sqrt{2}\big)A+\sqrt{4}I\big]$$

$$= \begin{bmatrix} 5-2\sqrt{2} & -1 & 2 \\ -1 & 4-2\sqrt{2} & -1 \\ -1 & 1 & 2-2\sqrt{2} \end{bmatrix}^{-1}\begin{bmatrix} 5-6\sqrt{2} & 2\sqrt{2}-1 & 2-4\sqrt{2} \\ 2\sqrt{2}-1 & 4-4\sqrt{2} & 2\sqrt{2}-1 \\ 2\sqrt{2}-1 & 1-2\sqrt{2} & 2 \end{bmatrix}$$

$$= \frac{3+2\sqrt{2}}{8}\begin{bmatrix} 3-2\sqrt{2} & 2\sqrt{2}+4 & -2\sqrt{2}-5 \\ 1 & 4-2\sqrt{2} & 1 \\ 1 & -2\sqrt{2}-4 & 9 \end{bmatrix}\begin{bmatrix} 5-6\sqrt{2} & 2\sqrt{2}-1 & 2-4\sqrt{2} \\ 2\sqrt{2}-1 & 4-4\sqrt{2} & 2\sqrt{2}-1 \\ 2\sqrt{2}-1 & 1-2\sqrt{2} & 2 \end{bmatrix}$$

$$= \frac{\sqrt{2}}{4}\begin{bmatrix} -5 & 4+2\sqrt{2} & -5-2\sqrt{2} \\ 1 & -4 & 1 \\ 1 & -4-2\sqrt{2} & 1+2\sqrt{2} \end{bmatrix}.$$

El lector puede verificar que esta matriz y su respectiva matriz negativa también son raíces cuadradas de A.

Teorema 3.26 *Si A es una matriz real de tamaño 4×4 con a lo más un valor propio nulo, entonces sus raíces cuadradas son:*

$$A^{\frac{1}{2}} = \big[\alpha A + \beta I\big]^{-1}\big[A^2 + \gamma A + \big(\sqrt{\lambda_1}\sqrt{\lambda_2}\sqrt{\lambda_3}\sqrt{\lambda_4}\big)I\big], \tag{3.29}$$

donde $\alpha = \sum_{k=1}^{4} \sqrt{\lambda_k}$, $\beta = \sum_{k>j>i} \sqrt{\lambda_i}\sqrt{\lambda_j}\sqrt{\lambda_k}$ y $\gamma = \sum_{j>i} \sqrt{\lambda_i}\sqrt{\lambda_j}$.

Demostración.

Supongamos que A es una matriz cuadrada con valores propios reales. Por el Teorema 3.7, es semejante a una matriz triangular superior T. Luego, puede expresarse como

$$A = QTQ^t, \tag{3.30}$$

donde Q es una matriz ortogonal y

$$T = \begin{bmatrix} \lambda_1 & a & b & c \\ 0 & \lambda_2 & d & e \\ 0 & 0 & \lambda_3 & f \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}. \tag{3.31}$$

Con $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ los valores propios de A. Al reemplazar (3.30) en (3.29), se tiene que

$$\begin{aligned} A^{\frac{1}{2}} &= \big[\alpha QTQ^t + \beta I\big]^{-1}\big[\big(QTQ^t\big)^2 + \gamma QTQ^t + \sqrt{\det\big(QTQ^t\big)}I\big] \\ &= \big[Q\big(\alpha T + \beta I\big)Q^t\big]^{-1}\big[Q\big(T^2 + \gamma T + \sqrt{\det T}I\big)Q^t\big] \\ &= Q\big[\alpha T + \beta I\big]^{-1}Q^tQ\big[T^2 + \gamma T + \sqrt{\det T}I\big]Q^t \\ &= Q\big[\alpha T + \beta I\big]^{-1}\big[T^2 + \gamma T + \sqrt{\det T}I\big]Q^t. \end{aligned}$$

Es decir, al utilizar la descomposición de Schur, se llega a que

$$A^{\frac{1}{2}} = QT^{\frac{1}{2}}Q^t.$$

Ahora, se demuestra que

$$T^{\frac{1}{2}} = \big[\alpha T + \beta I\big]^{-1}\big[T^2 + \gamma T + \sqrt{\det T}I\big].$$

Usando la expresión (3.26), se calculan $\alpha,\ \beta$ y γ, como sigue

$$\begin{aligned}\alpha &= \lambda_{12}+\lambda_{34},\\ \beta &= \sqrt{\lambda_1}\sqrt{\lambda_2}\lambda_{34}+\sqrt{\lambda_3}\sqrt{\lambda_4}\lambda_{12},\\ \gamma &= \sqrt{\lambda_1}\lambda_{23}+\sqrt{\lambda_2}\lambda_{34}+\sqrt{\lambda_4}\lambda_{13}.\end{aligned}$$

Luego, se tiene que

$$\alpha A+\beta I=\begin{bmatrix}\omega_1 & a\alpha & b\alpha & c\alpha\\ 0 & \omega_2 & d\alpha & e\alpha\\ 0 & 0 & \omega_3 & f\alpha\\ 0 & 0 & 0 & \omega_4\end{bmatrix},$$

donde

$$\omega_i = \prod_{j\neq i}^{4}\lambda_{ij}, \qquad\qquad i=1,2,3,4,$$

y la inversa es

$$\big(\alpha A+\beta I\big)^{-1}=\frac{1}{\xi}\begin{bmatrix}\tau_1 & -a\alpha\frac{\lambda_{34}}{\lambda_{12}} & \alpha\frac{\alpha ad-\omega_2 b}{\tau_4} & -\alpha\frac{\alpha^2 adf-\alpha\omega_3 ae-\alpha\omega_2 bf+\omega_2\omega_3 c}{\xi}\\ 0 & \tau_2 & -d\alpha\frac{\lambda_{14}}{\lambda_{23}} & \alpha\frac{d\alpha f-e\omega_3}{\tau_1}\\ 0 & 0 & \tau_3 & -f\alpha\frac{\lambda_{12}}{\lambda_{34}}\\ 0 & 0 & 0 & \tau_4\end{bmatrix},$$

donde

$$\begin{aligned}\tau_1=\omega_2\frac{\lambda_{34}}{\lambda_{12}}=\omega_3\frac{\lambda_{24}}{\lambda_{13}}=\omega_4\frac{\lambda_{23}}{\lambda_{14}}, &\qquad \tau_2=\omega_1\frac{\lambda_{34}}{\lambda_{12}}=\omega_3\frac{\lambda_{14}}{\lambda_{23}}=\omega_4\frac{\lambda_{13}}{\lambda_{24}},\\ \tau_3=\omega_1\frac{\lambda_{24}}{\lambda_{13}}=\omega_2\frac{\lambda_{14}}{\lambda_{23}}=\omega_4\frac{\lambda_{12}}{\lambda_{34}}, &\qquad \tau_4=\omega_1\frac{\lambda_{23}}{\lambda_{14}}=\omega_2\frac{\lambda_{13}}{\lambda_{24}}=\omega_3\frac{\lambda_{12}}{\lambda_{34}},\end{aligned}$$
$$\xi=\omega_1\tau_1=\omega_2\tau_2=\omega_3\tau_3=\omega_4\tau_4.$$

Por otra parte,

$$T^2=\begin{bmatrix}\lambda_1^2 & a\big(\lambda_1+\lambda_2\big) & b\big(\lambda_1+\lambda_3\big)+ad & c\big(\lambda_1+\lambda_4\big)+ae+bf\\ 0 & \lambda_2^2 & d\big(\lambda_2+\lambda_3\big) & e\big(\lambda_2+\lambda_4\big)+df\\ 0 & 0 & \lambda_3^2 & f\big(\lambda_3+\lambda_4\big)\\ 0 & 0 & 0 & \lambda_4^2\end{bmatrix},$$

por lo tanto, la matriz $T^2 + \gamma T + \sqrt{\det(T)}I$ es igual a

$$\begin{bmatrix} \omega_1\sqrt{\lambda_1} & a\alpha\sqrt{\lambda_2} + \frac{a\omega_1}{\lambda_{12}} & b\alpha\sqrt{\lambda_3} + \frac{b\omega_1}{\lambda_{13}} + ad & c\alpha\sqrt{\lambda_4} + \frac{c\omega_1}{\lambda_{14}} + ae + bf \\ 0 & \omega_2\sqrt{\lambda_2} & d\alpha\sqrt{\lambda_3} + \frac{d\omega_2}{\lambda_{23}} & e\alpha\sqrt{\lambda_4} + \frac{e\omega_2}{\lambda_{24}} + df \\ 0 & 0 & \omega_3\sqrt{\lambda_3} & f\alpha\sqrt{\lambda_4} + \frac{f\omega_3}{\lambda_{34}} \\ 0 & 0 & 0 & \omega_4\sqrt{\lambda_4} \end{bmatrix}.$$

Al realizar los respectivos productos, se llega a

$$T^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & \frac{a}{\lambda_{12}} & \frac{b}{\lambda_{13}} - \frac{ad}{\tau_4} & \frac{c}{\lambda_{14}} - \frac{bf}{\tau_2} - \frac{ae}{\tau_3} + adf\frac{\alpha}{\xi} \\ 0 & \sqrt{\lambda_2} & \frac{d}{\lambda_{23}} & \frac{e}{\lambda_{24}} - \frac{df}{\tau_1} \\ 0 & 0 & \sqrt{\lambda_3} & \frac{f}{\lambda_{34}} \\ 0 & 0 & 0 & \sqrt{\lambda_4} \end{bmatrix}.$$

El lector puede verificar que los elementos de esta última matriz coinciden con los dados en el Teorema 3.23. ■

Ejemplo 3.16 *Determine, mediante el método descrito en el teorema anterior, una raíz cuadrada para la matriz dada en el Ejemplo 3.13.*

Solución.

Como en el Ejemplo 3.13 se obtuvieron los valores propios de A, se tiene que

$$\alpha = \sqrt{4} + \sqrt{9} + \sqrt{16} + \sqrt{25} = 14,$$
$$\beta = \sqrt{4}\sqrt{9}\sqrt{16} + \sqrt{4}\sqrt{9}\sqrt{25} + \sqrt{4}\sqrt{16}\sqrt{25} + \sqrt{9}\sqrt{16}\sqrt{25} = 154,$$
$$\gamma = \sqrt{4}\sqrt{9} + \sqrt{4}\sqrt{16} + \sqrt{4}\sqrt{25} + \sqrt{9}\sqrt{16} + \sqrt{9}\sqrt{25} + \sqrt{16}\sqrt{25} = 71.$$

Por lo tanto, la matriz $\alpha A + \beta I$ es

$$\begin{bmatrix} 322 & 14 & -28 & -154 \\ 98 & 308 & -70 & 168 \\ -140 & -42 & 378 & -14 \\ -42 & 56 & 98 & 364 \end{bmatrix},$$

y por otra parte, la matriz $A^2 + \gamma A + \sqrt{\det A}I$ es

$$\begin{bmatrix} 1176 & 56 & -280 & -1064 \\ 672 & 1092 & -420 & 1092 \\ -1008 & -308 & 1540 & -28 \\ -336 & 364 & 700 & 1484 \end{bmatrix}.$$

Luego, la raíz cuadrada de A es

$$A^{\frac{1}{2}} = \frac{1}{22680}\begin{bmatrix} 76 & -11 & -6 & 37 \\ -21 & 87 & 27 & -48 \\ 26 & 5 & 60 & 11 \\ 5 & -16 & -21 & 71 \end{bmatrix}\begin{bmatrix} 1176 & 56 & -280 & -1064 \\ 672 & 1092 & -420 & 1092 \\ -1008 & -308 & 1540 & -28 \\ -336 & 364 & 700 & 1484 \end{bmatrix}$$
$$= \frac{1}{3}\begin{bmatrix} 10 & 1 & 0 & -5 \\ 3 & 9 & -3 & 6 \\ -4 & -1 & 12 & -1 \\ -1 & 2 & 3 & 11 \end{bmatrix},$$

la cual coincide con una de las obtenidas en el Ejemplo 3.13.

Teorema 3.27 *Si A es una matriz real de tamaño $n \times n$ ($n \geq 5$), con una descomposición de la forma $A = PBP^{-1}$, entonces sus raíces cuadradas se calculan de la siguiente manera:*

$$A^{\frac{1}{2}} = PB^{\frac{1}{2}}P^{-1} = P\begin{bmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_k \end{bmatrix}^{\frac{1}{2}} P^{-1}, \tag{3.32}$$

en donde cada submatriz B_m es de tamaño 1×1, 2×2, 3×3 o 4×4, de tal manera que se le pueda calcular a cada bloque una raíz cuadrada como las dadas en (3.17), (3.22) o (3.29), respectivamente.

Demostración.

Queda como ejercicio para el lector. ■

Ejemplo 3.17 *Determine una raíz cuadrada para la siguiente matriz:*

$$A = \begin{bmatrix} 8 & 4 & -2 \\ -7 & -1 & 3 \\ 1 & 3 & 1 \end{bmatrix}.$$

Solución.

Para la matriz A, se tiene que el polinomio característico es

$$p_A(\lambda) = -\lambda^3 + 8\lambda^2 - 20\lambda,$$

luego se tiene que

$$\lambda_1 = 0, \qquad \lambda_2 = 4 + 2i, \qquad \lambda_3 = 4 - 2i.$$

Como A tiene valores propios complejos, usando el método de factorización dado en (2.22), se tiene

$$\begin{bmatrix} 8 & 4 & -2 \\ -7 & -1 & 3 \\ 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 4 & 2 & 0 \\ -2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix}^{-1}.$$

Luego, la raíz cuadrada de A es

$$\begin{aligned} A^{\frac{1}{2}} &= \frac{1}{2} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 4 & 2 & 0 \\ -2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} 2 & 2 & 0 \\ -3 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} 4 & 2 \\ -2 & 4 \end{bmatrix}^{\frac{1}{2}} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 & 0 \\ -3 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}. \end{aligned}$$

Si se usa una de las raíces cuadradas encontradas en el Ejemplo 3.12, se tiene que

$$\begin{aligned} A^{\frac{1}{2}} &= \frac{1}{2\Delta} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} 4+2\sqrt{5} & 2 \\ -2 & 4+2\sqrt{5} \end{bmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 & 0 \\ -3 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 2\sqrt{5}+1 & 2\sqrt{5}+3 & 1 \\ -8-3\sqrt{5} & -4-\sqrt{5} & \sqrt{5}+2 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} 8+3\sqrt{5} & 4+\sqrt{5} & -2-\sqrt{5} \\ -7-\sqrt{5} & \sqrt{5}-1 & 3+\sqrt{5} \\ 1+2\sqrt{5} & 3+2\sqrt{5} & 1 \end{bmatrix}, \end{aligned}$$

donde $\Delta = 2\sqrt{2+\sqrt{5}}$. Se puede fácilmente verificar que $\left(A^{\frac{1}{2}}\right)^2 = A$

$$\begin{aligned}\left(A^{\frac{1}{2}}\right)^2 &= \left(\frac{1}{2\sqrt{2+\sqrt{5}}}\right)^2 \begin{bmatrix} 8+3\sqrt{5} & 4+\sqrt{5} & -2-\sqrt{5} \\ -7-\sqrt{5} & \sqrt{5}-1 & 3+\sqrt{5} \\ 1+2\sqrt{5} & 3+2\sqrt{5} & 1 \end{bmatrix}^2 \\ &= \frac{1}{4\left(2+\sqrt{5}\right)} \begin{bmatrix} 64+32\sqrt{5} & 32+16\sqrt{5} & -16-8\sqrt{5} \\ -56-28\sqrt{5} & -8-4\sqrt{5} & 24+12\sqrt{5} \\ 8+4\sqrt{5} & 24+12\sqrt{5} & 8+4\sqrt{5} \end{bmatrix} \\ &= \begin{bmatrix} 8 & 4 & -2 \\ -7 & -1 & 3 \\ 1 & 3 & 1 \end{bmatrix}.\end{aligned}$$

3.5.1 Raíces cuadradas de matrices simétricas

De manera análoga a la Sección 2.4, en la cual se desarrolló el tema sobre diagonalización para *matrices simétricas*, en este apartado se presenta por separado la parte concerniente a raíces cuadradas para matrices simétricas.

Teorema 3.28 *Toda matriz simétrica A de tamaño $n \times n$ tiene valores propios positivos si y solo si existe una matriz simétrica B de tamaño $n \times n$ tal que*

$$A = B^2. \tag{3.33}$$

La matriz B se denomina una raíz cuadrada de A.

Demostración.

Si los valores propios de A son positivos, entonces $\det(A) > 0$, y por ser simétrica, se puede factorizar de la forma

$$\begin{aligned}A &= QDQ^t = Q\left(D^{\frac{1}{2}}\right)^2 Q^t \\ &= \left(QD^{\frac{1}{2}}Q^t\right)\left(QD^{\frac{1}{2}}Q^t\right) \\ &= B^tB.\end{aligned}$$

Nótese que B es una matriz simétrica y de rango n (como A), por lo tanto $B^tB = B^2$. La matriz $D^{\frac{1}{2}} = diag\left\{\sqrt{\lambda_i}\right\}$ semejante a B está definida como

en (3.18). Además, si se consideran únicamente las raíces positivas de los valores propios, la matriz B tendrá también valores propios positivos (como A). ■

Ejemplo 3.18 *Determine una raíz cuadrada para la matriz dada en el Ejemplo 2.18.*

Solución.

Haciendo referencia al Ejemplo 2.18, se tiene que

$$Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \quad \text{y} \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}.$$

Por lo tanto,

$$\begin{aligned} B &= \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sqrt{7} \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \\ &= \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ \sqrt{7}/\sqrt{6} & 2\sqrt{7}/\sqrt{6} & \sqrt{7}/\sqrt{6} \end{bmatrix}. \end{aligned}$$

Es decir,

$$B = \frac{1}{6} \begin{bmatrix} 5+\sqrt{7} & -2+2\sqrt{7} & -1+\sqrt{7} \\ -2+2\sqrt{7} & 2+4\sqrt{7} & -2+2\sqrt{7} \\ -1+\sqrt{7} & -2+2\sqrt{7} & 5+\sqrt{7} \end{bmatrix}.$$

El lector puede verificar que $A = B^2$.

De acuerdo con la Definición 2.7, las matrices reales cuadradas A y B se dicen que son *congruentes* si existe una matriz P no singular tal que

$$A = P^t B P.$$

En ocasiones, además de establecer el hecho en sí de la congruencia, se requiere encontrar la matriz P de la transformación que satisface $A = P^t B P$. En estos momentos, se puede construir la matriz P para matrices simétricas no singulares A y B utilizando la descomposición LDU de cada matriz y una de las raíces cuadradas de D, como sigue

$$P^t = L_A D_A^{\frac{1}{2}} D_B^{-\frac{1}{2}} L_B^{-1}, \tag{3.34}$$

donde las descomposiciones LDU para las matrices A y B son $L_A D_A L_A^t$ y $L_B D_B L_B^t$, respectivamente.

Ejemplo 3.19 *Determine si las matrices dadas en los Ejemplos 2.18 y 3.3 son congruentes.*

Solución.

La factorización LDL^t de la matriz dada en el Ejemplo 2.18 es

$$\begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \frac{7}{6} \end{bmatrix} \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} = L_A D_A L_A^t.$$

En el Ejemplo 3.3, se obtuvo que

$$\begin{bmatrix} 1 & 3 & 5 \\ 3 & 12 & 18 \\ 5 & 18 & 30 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = L_B D_B L_B^t.$$

Por lo tanto,

$$\begin{aligned} P^t &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{\frac{7}{6}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -2 & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{2} & 0 & 0 \\ \sqrt{2} - 3 & 1 & 0 \\ \frac{1}{2}\sqrt{2} - \frac{1}{3}\sqrt{21} - 1 & \frac{1}{3} - \frac{1}{6}\sqrt{21} & \frac{1}{6}\sqrt{21} \end{bmatrix}. \end{aligned}$$

El lector puede verificar que $A = P^t B P$.

Teorema 3.29 *Una matriz simétrica A de tamaño $n \times n$ tiene todos sus valores propios positivos ($\lambda_i > 0$) si y solo si:*

$$A = P^t P, \tag{3.35}$$

donde P es no singular.

Demostración.

Si A es simétrica y todos sus valores propios son positivos, entonces puede escribirse en la forma

$$A = QDQ^t = \left(QD^{\frac{1}{2}}\right)\left(D^{\frac{1}{2}}Q^t\right) = P^t P,$$

con $P = D^{\frac{1}{2}}Q^t$ y $D^{\frac{1}{2}} = diag\left\{\sqrt{\lambda_i}\right\}$ definida como en (3.18). ■

3.5.2 Descomposición de Cholesky

Entre los tipos de factorizaciones para la matriz A, existe una descomposición especial para aquellas matrices cuadradas cuyos valores propios son todos positivos conocida como *descomposición de Cholesky*, considerada en este apartado.

Teorema 3.30 Descomposición de Cholesky

Si A es una matriz simétricade tamaño $n \times n$ con todos sus valores propios positivos, entonces existe una matriz L triangular inferior tal que

$$A = LL^t, \tag{3.36}$$

donde todos los elementos en la diagonal principal de L son positivos.

Demostración.

Por el Teorema 3.5, la matriz A se puede expresar como

$$A = LDL^t = \left(LD^{\frac{1}{2}}\right)\left(D^{\frac{1}{2}}L^t\right) = \left(LD^{\frac{1}{2}}\right)\left(LD^{\frac{1}{2}}\right)^t,$$

donde $D^{\frac{1}{2}}$ está definida como en (3.18) y la prueba queda completa. ■

A continuación, se presenta un procedimiento para encontrar la descomposición de Cholesky.

Procedimiento para encontrar los elementos de $R = D^{\frac{1}{2}}L^t$

- Para $i = 1$, se tiene

$$r_{1j} = \begin{cases} \sqrt{a_{11}} & j = 1, \\ r_{11}^{-1}a_{1j} & j > 1. \end{cases}$$

- Cuando $i > 1$, se obtiene

$$r_{ij} = \begin{cases} 0 & i > j, \\ \sqrt{a_{ii} - \sum_{k=1}^{i-1} r_{ki}^2} & i = j, \\ r_{ii}^{-1}\left(a_{ij} - \sum_{k=1}^{i-1} r_{ki}r_{kj}\right) & i < j. \end{cases}$$

El procedimiento exige que estos elementos se calculen por filas, de izquierda a derecha y de arriba para abajo.

Si no se puede obtener la descomposición de Cholesky de una matriz (por ejemplo, cuando al realizar el procedimiento presentado arriba surge una raíz cuadrada de un número negativo), esto es indicio de que la matriz simétrica no tiene todos sus valores propios positivos.

Ejemplo 3.20 *Encuentre la descomposición de Cholesky para la matriz simétrica dada en el Ejemplo 3.3.*

Solución.

Usando el procedimiento descrito anteriormente, se tiene que

$$\begin{aligned} r_{11} &= \sqrt{a_{11}} = 1, & r_{12} &= r_{11}^{-1} a_{12} = a_{12} = 3, \\ r_{13} &= r_{11}^{-1} a_{13} = a_{13} = 5, & r_{22} &= \sqrt{a_{22} - r_{12}^2} = \sqrt{3}, \\ r_{23} &= r_{22}^{-1}\big(a_{23} - r_{12} r_{13}\big) = \sqrt{3}, & r_{33} &= \sqrt{a_{33} - r_{13}^2 - r_{23}^2} = \sqrt{2}. \end{aligned}$$

Luego,

$$R = \begin{bmatrix} 1 & 3 & 5 \\ 0 & \sqrt{3} & \sqrt{3} \\ 0 & 0 & \sqrt{2} \end{bmatrix}.$$

El lector puede verificar que $A = R^t R$.

Ejercicios 3.3

1. Si A es una matriz real de tamaño $n \times n$ con valores propios distintos $\lambda_1, \lambda_2, \ldots, \lambda_n$, usando la descomposición de Sylvester pruebe que la m-ésima potencia de A está dada por

$$A^m = \sum_{i=1}^{n} \lambda_i^m \mathbf{E}\big(\lambda_i\big), \qquad m \in \mathbb{Z}; \tag{3.37}$$

cuando $m \in \mathbb{Q}$, este resultado también se cumple.

2. Para cada una de las siguientes matrices, determine (en caso de ser posible) una raíz cuadrada:

$$a.\begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}. \quad b.\begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}. \quad c.\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}. \quad d.\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}.$$

$$e.\begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}. \quad f.\begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}. \quad g.\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}. \quad h.\begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$$

3. Determine la descomposición de Cholesky para las siguientes matrices:

$$a.\begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}. \quad b.\begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}. \quad c.\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

$$d.\begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -2 & 2 & 1 \end{bmatrix}. \quad e.\begin{bmatrix} 3 & -1 & 4 \\ -1 & 1 & -5 \\ 4 & -5 & 2 \end{bmatrix}.$$

4. Determine las soluciones de las siguientes ecuaciones matriciales:

$a.\ X^2 - 3X = A$, con

$$A = \begin{bmatrix} -3 & 3 & 6 \\ -\frac{7}{3} & 5 & 2 \\ -\frac{4}{3} & 4 & 0 \end{bmatrix}$$

$b.\ X^2 - 4X - 3I = B$, con

$$B = \begin{bmatrix} 7 & -2 & 4 \\ -2 & 7 & 4 \\ 4 & 4 & 1 \end{bmatrix}.$$

3.6 Descomposición en valores singulares

Una factorización especial para cualquier matriz A de tamaño $m \times n$ es la *descomposición en valores singulares* (SVD, por su sigla en inglés), la

cual es una de las factorizaciones de matrices más útiles en álgebra lineal aplicada.

Definición 3.10 Valores singulares de matrices cuadradas

Los valores singulares de una matriz real A de tamaño $n \times n$ son las raíces cuadradas de los valores propios asociados a la matriz simétrica A^tA (listados con sus multiplicidades algebraicas). Estos valores se denotan por $\sigma_1, \sigma_2, \ldots, \sigma_n$, y se colocan en orden decreciente:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0,$$

donde $\sigma_i = \sqrt{\lambda_i}$ para $1 \leq i \leq n$.

Ejemplo 3.21 *Determine los valores singulares de la matriz:*

$$A = \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$$

Solución.

La matriz A^tA es

$$\begin{bmatrix} 1 & 3 & -1 \\ 3 & 1 & 1 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 5 & 3 \\ 5 & 11 & -3 \\ 3 & -3 & 9 \end{bmatrix}.$$

En este caso, la ecuación característica es

$$\det\big(A^tA - \lambda I\big) = -\lambda^3 + 31\lambda^2 - 276\lambda + 576 = 0.$$

Entonces, los valores propios de A^tA son $\lambda_1 = 16, \lambda_2 = 12$ y $\lambda_3 = 3$. Por lo tanto, los valores singulares de la matriz A son $\sigma_1 = 4, \sigma_2 = 2\sqrt{3}$ y $\sigma_3 = \sqrt{3}$.

Teorema 3.31 *Si A es una matriz simétrica, los absolutos de los valores propios son los valores singulares.*

Demostración.

Como $A = A^t$ todos sus valores propios $\lambda_i \in \mathbb{R}$, entonces

$$\begin{aligned} \det\big(A^tA - \sigma^2 I\big) &= \det\big(A^2 - \sigma^2 I\big) \\ &= \det\big(A - |\sigma| I\big)\det\big(A + |\sigma| I\big), \end{aligned}$$

luego

$$\sigma_i = \lambda_i \qquad \text{o} \qquad \sigma_i = -\lambda_i. \ \blacksquare$$

Cuando A es una matriz real de tamaño $n \times n$, sabemos que las matrices A^tA y AA^t tienen los mismos valores propios con las mismas multiplicidades algebraicas. Por lo tanto, en la Definición 3.10 se puede cambiar A^tA por AA^t. Mientras que si A es una matriz real de tamaño $m \times n$, con $m \neq n$, las matrices A^tA y AA^t tendrán n y m valores propios, respectivamente. Por consiguiente, cuando la matriz no sea cuadrada, sus valores singulares se definen de la siguiente manera.

Definición 3.11 Valores singulares de matrices rectangulares

Sea A una matriz real de tamaño $m \times n$ ($m \neq n$), los valores singulares son las raíces cuadradas de los valores propios comunes a las matrices simétricas A^tA y AA^t.

Ejemplo 3.22 *Encuentre los valores singulares de la matriz:*

$$A = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 1 & 1 \end{bmatrix}.$$

Solución.

La matriz AA^t es

$$\begin{bmatrix} 1 & 3 & -2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 3 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 0 \\ 0 & 3 \end{bmatrix},$$

cuyos valores propios son $\lambda_1 = 14$ y $\lambda_2 = 3$. La matriz A^tA es

$$\begin{bmatrix} 1 & -1 \\ 3 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & -3 \\ 2 & 10 & -5 \\ -3 & -5 & 5 \end{bmatrix}.$$

En este caso, los valores propios de A^tA son $\lambda_1 = 14, \lambda_2 = 3$ y $\lambda_3 = 0$. Por lo tanto, los valores singulares de la matriz A son $\sigma_1 = \sqrt{14}$ y $\sigma_2 = \sqrt{3}$.

Teorema 3.32 *Sea A una matriz real de tamaño $m \times n$ que tiene r valores singulares no nulos $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ con $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$, entonces el rango de A es r.*

Demostración.

Sea $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ una base ortonormal de $\mathbb{R}^n$ formada por los vectores propios asociados a A^tA, y ordenados de tal forma que los valores propios correspondientes a A^tA satisfacen que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Entonces,

$$\begin{aligned} A\vec{v}_i \cdot A\vec{v}_j &= \big(A\vec{v}_i\big)^t A\vec{v}_j = \vec{v}_i^t\big(A^tA\vec{v}_j\big) = \vec{v}_i^t\lambda_j\vec{v}_j \\ &= \sigma_j^2\big(\vec{v}_i \cdot \vec{v}_j\big) = \sigma_j^2\delta_{ij} = \begin{cases} \sigma_j^2 & \text{si } \ i = j, \\ 0 & \text{si } \ i \neq j. \end{cases} \end{aligned} \tag{3.38}$$

Luego, $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_n\}$ es un conjunto ortogonal. Sea r el número de valores singulares no nulos de A, esto es, r es el número de valores propios no nulos de A^tA. De la expresión (3.38), se tiene que $A\vec{v}_i \neq \vec{0}$ si y solo si $1 \leq i \leq r$. Entonces, $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$ son vectores linealmente independientes, los cuales claramente pertenecen al espacio columna de A $\big[Col(A)\big]$. Además, para cualquier $\vec{y} \in Col(A)$ −digamos, $\vec{y} = A\vec{x}$− se puede escribir $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n$, y

$$\begin{aligned} \vec{y} = A\vec{x} &= c_1A\vec{v}_1 + \ldots + c_rA\vec{v}_r + c_{r+1}A\vec{v}_{r+1} + \ldots + c_nA\vec{v}_n \\ &= c_1A\vec{v}_1 + \ldots + c_rA\vec{v}_r + \vec{0} + \ldots + \vec{0}. \end{aligned}$$

Así que $\vec{y}$ está en el espacio generado por $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$, lo cual muestra que $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$ es una base (ortogonal) para $Col(A)$. Por lo tanto, el $\rho(A) = r$. ■

3.6.1 Descomposición en valores singulares

La descomposición de A involucra una matriz "diagonal" S de tamaño $m \times n$ particionada como sigue

$$S = \begin{bmatrix} D_{r\times r} & \vdots & \mathrm{O}_{r\times n_1} \\ \ldots & \cdot & \ldots \\ \mathrm{O}_{m_1\times r} & \vdots & \mathrm{O}_{m_1\times n_1} \end{bmatrix} \quad \text{con} \quad D_{r\times r} = \begin{bmatrix} \sigma_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_r \end{bmatrix}, \tag{3.39}$$

donde $m_1 = m - r$, $n_1 = n - r$, σ_i, para $i = 1, 2, \ldots, r$ son los valores singulares no nulos de A y $r = \rho(A)$. (Si r es igual a m o n, entonces algunas o todas las matrices nulas desaparecen).

Teorema 3.33 Descomposición en valores singulares

Sea A una matriz realde tamaño $m \times n$ con rango r. Entonces, existen matrices ortogonales U y V de tamaño $m \times m$ y $n \times n$, respectivamente, tales que

$$A = USV^t, \tag{3.40}$$

donde S tiene la forma dada en la expresión (3.39).

Demostración.

Sean λ_i y $\vec{v}_i$ como en la prueba del Teorema 3.32. Entonces, $\sigma_i = \sqrt{\lambda_i} = \|A\vec{v}_i\| > 0$ para $1 \leq i \leq r$, $r = \rho(A) \leq \text{mín}\{m, n\}$, y el conjunto de vectores $\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$ es una base ortogonal para $Col(A)$. Si se normalizan cada uno de los vectores $A\vec{v}_i$, se puede definir

$$\vec{u}_i = \frac{1}{\|A\vec{v}_i\|} A\vec{v}_i = \frac{1}{\sigma_i} A\vec{v}_i, \qquad i = 1, 2, \ldots, r.$$

Luego, el conjunto de vectores $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$ es una base ortonormal para $Col(A)$; esta base se puede extender hasta obtenerse una base ortonormal de $\mathbb{R}^m$: expresémosla por $\{\vec{u}_1, \ldots, \vec{u}_r, \vec{u}_{r+1}, \ldots, \vec{u}_m\}$. A partir de la definición de los vectores $\vec{u}_i$, se puede escribir

$$A\vec{v}_i = \begin{cases} \sigma_i \vec{u}_i & \text{para } i = 1, 2, \ldots, r \\ 0\vec{u}_i & \text{para } i = r+1, r+2, \ldots, m \\ \vec{0} & \text{para } i = m+1, m+2, \ldots, n. \end{cases}$$

En forma matricial se expresa de la siguiente manera

$$\begin{aligned} AV &= \begin{bmatrix} A\vec{v}_1 & \ldots & A\vec{v}_r & A\vec{v}_{r+1} & \ldots & A\vec{v}_m & A\vec{v}_{m+1} & \ldots & A\vec{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1\vec{u}_1 & \ldots & \sigma_r\vec{u}_r & 0\vec{u}_{r+1} & \ldots & 0\vec{u}_m & \vec{0} & \ldots & \vec{0} \end{bmatrix} \\ &= \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_r & \vec{u}_{r+1} & \ldots & \vec{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_r & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 \end{bmatrix}. \end{aligned}$$

Por lo tanto,

$$AV = \begin{bmatrix} \vec{u}_1 & \dots & \vec{u}_r & \vec{u}_{r+1} & \dots & \vec{u}_m \end{bmatrix} \begin{bmatrix} D_{r\times r} & \vdots & \mathrm{O}_{r\times n_1} \\ \dots & \cdot & \dots \\ \mathrm{O}_{m_1\times r} & \vdots & \mathrm{O}_{m_1\times n_1} \end{bmatrix} = US. \tag{3.41}$$

Nótese que las columnas de la matriz ortogonal V (de tamaño $n \times n$) son los vectores propios ortonormalizados de la matriz A^tA. Por otra parte, las columnas de la matriz ortogonal U (de tamaño $m \times m$) son los vectores propios ortonormalizados de la matriz AA^t y la matriz S está definida como en (3.39). Si se multiplica por el lado derecho de la ecuación (3.41) por V^{-1} ($V^{-1} = V^t$), se tiene que

$$A = USV^t.$$

Esto finaliza la demostración del teorema. ■

Ejemplo 3.23 *Encuentre la descomposición en valores singulares de la matriz dada en el Ejemplo 3.21.*

Solución.

Del Ejemplo 3.21 se tiene que los valores singulares asociados a la matriz A son $\sigma_1^2 = 16$, $\sigma_2^2 = 12$ y $\sigma_3^2 = 3$. Al calcular los respectivos vectores propios ortonormalizados de A^tA, se obtiene:

$$\vec{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \quad \text{y} \quad \vec{v}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

Por otra parte, la matriz AA^t es

$$AA^t = \begin{bmatrix} 14 & 2 & 0 \\ 2 & 14 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

y sus respectivos vectores propios ortonormalizados son

$$\vec{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{y} \quad \vec{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Finalmente, si $U = [\vec{u}_1\ \vec{u}_2\ \vec{u}_3]$, $V = [\vec{v}_1\ \vec{v}_2\ \vec{v}_3]$ y $S = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}$, entonces

$$A = \begin{bmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & 0 \\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 \\ \frac{1}{6}\sqrt{6} & -\frac{1}{6}\sqrt{6} & \frac{1}{3}\sqrt{6} \\ -\frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{3} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$$

Ejemplo 3.24 *Encuentre la descomposición en valores singulares de la matriz dada en el Ejemplo 3.22.*

Solución.

Haciendo referencia al Ejemplo 3.22, se tiene que los valores singulares asociados a la matriz A son $\sigma_1^2 = 14$ y $\sigma_3^2 = 3$. Al calcular los respectivos vectores propios ortonormalizados de AA^t, se obtiene:

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Por otra parte, los respectivos vectores propios ortonormalizados de A^tA son

$$\vec{v}_1 = \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \quad \vec{v}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \quad \text{y} \quad \vec{v}_3 = \frac{1}{\sqrt{42}} \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}.$$

Si $U = [\vec{u}_1\ \vec{u}_2]$, $V = [\vec{v}_1\ \vec{v}_2\ \vec{v}_3]$ y $S = \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{bmatrix}$, entonces

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{14}} & -\frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{42}} \\ -\frac{2}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{4}{\sqrt{42}} \end{bmatrix} = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 1 & 1 \end{bmatrix}.$$

3.6.2 Descomposición polar

Una consecuencia interesante y útil de la descomposición en valores singulares para una matriz *cuadrada* A es la *descomposición polar* de A.

Teorema 3.34 Descomposición polar a izquierda

Sea A una matriz real de tamaño $n \times n$ con rango r. Entonces, existe una matriz simétrica P de tamaño $n \times n$ con valores propios no negativos y una matriz ortogonal Q de tamaño $n \times n$ tales que

$$A = PQ. \tag{3.42}$$

Demostración.

Si A es una matriz real de tamaño $n \times n$, en una descomposición de valores singulares las matrices U, S y V son también de tamaño $n \times n$. En este caso, se puede escribir la ecuación (3.40) como

$$A = USV^t = US(U^tU)V^t = (USU^t)UV^t = PQ,$$

la matriz P es siempre única, aun cuando A sea singular, y está dada por

$$P = \left[AA^t\right]^{\frac{1}{2}} = USU^t;$$

nótese que esta matriz es simétrica y $Q = UV^t$ es una matriz ortogonal. Se deja como ejercicio la comprobación de que P tiene valores propios no negativos. ■

Corolario 3.34.1 Descomposición polar a derecha

Sea A una matriz compleja de tamaño $n \times n$ con rango r. Entonces, existe una matriz simétrica P' de tamaño $n \times n$ con valores propios no negativos y una matriz ortogonal Q de tamaño $n \times n$ tales que

$$A = QP'. \tag{3.43}$$

Demostración.

En términos de la descomposición en valores singulares de A, es decir, $A = USV^t$, se tiene que

$$P' = \left[A^t A\right]^{\frac{1}{2}} = VSV^t \qquad \text{y} \qquad Q = UV^t.$$

Si A es no singular, entonces la matriz Q está dada por

$$Q = A\left[P'\right]^{-1}.$$

El lector puede verificar que P' tiene valores propios no negativos y Q es ortogonal. ■

Ejemplo 3.25 *Encuentre la descomposición polar tanto a izquierda como a derecha para la matriz dada en el Ejemplo 3.21.*

Solución.

En el Ejemplo 3.23, se obtuvo la descomposición de A en valores singulares mediante las matrices U, S y V:

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}, \qquad S = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}$$

y

$$V = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & 1 & -\sqrt{2} \\ \sqrt{3} & -1 & \sqrt{2} \\ 0 & 2 & \sqrt{2} \end{bmatrix}.$$

Si se definen

$$P = USU^t = \left(\frac{1}{\sqrt{2}}\right)^2 \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 2+\sqrt{3} & 2-\sqrt{3} & 0 \\ 2-\sqrt{3} & 2+\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}$$

y

$$Q = UV^t = \frac{1}{2\sqrt{3}} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{3} & \sqrt{3} & 0 \\ 1 & -1 & 2 \\ -\sqrt{2} & \sqrt{2} & \sqrt{2} \end{bmatrix},$$

simplificando

$$Q = \frac{\sqrt{3}}{6}\begin{bmatrix} \sqrt{3}-1 & \sqrt{3}+1 & -2 \\ \sqrt{3}+1 & \sqrt{3}-1 & 2 \\ -2 & 2 & 2 \end{bmatrix},$$

entonces la descomposición polar a izquierda queda

$$\begin{aligned} A &= \frac{\sqrt{3}}{6}\begin{bmatrix} 2+\sqrt{3} & 2-\sqrt{3} & 0 \\ 2-\sqrt{3} & 2+\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}\begin{bmatrix} \sqrt{3}-1 & \sqrt{3}+1 & -2 \\ \sqrt{3}+1 & \sqrt{3}-1 & 2 \\ -2 & 2 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}. \end{aligned}$$

Para determinar la descomposición polar a derecha, se establece

$$\begin{aligned} P' &= \left(\frac{1}{\sqrt{6}}\right)^2\begin{bmatrix} \sqrt{3} & 1 & -\sqrt{2} \\ \sqrt{3} & -1 & \sqrt{2} \\ 0 & 2 & \sqrt{2} \end{bmatrix}\begin{bmatrix} 4 & 0 & 0 \\ 0 & 2\sqrt{3} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}\begin{bmatrix} \sqrt{3} & \sqrt{3} & 0 \\ 1 & -1 & 2 \\ -\sqrt{2} & \sqrt{2} & \sqrt{2} \end{bmatrix} \\ &= \frac{\sqrt{3}}{3}\begin{bmatrix} 2\sqrt{3}+2 & 2\sqrt{3}-2 & 1 \\ 2\sqrt{3}-2 & 2\sqrt{3}+2 & -1 \\ 1 & -1 & 5 \end{bmatrix}, \end{aligned}$$

luego la descomposición polar a derecha queda

$$\begin{aligned} A &= \frac{1}{6}\begin{bmatrix} \sqrt{3}-1 & \sqrt{3}+1 & -2 \\ \sqrt{3}+1 & \sqrt{3}-1 & 2 \\ -2 & 2 & 2 \end{bmatrix}\begin{bmatrix} 2\sqrt{3}+2 & 2\sqrt{3}-2 & 1 \\ 2\sqrt{3}-2 & 2\sqrt{3}+2 & -1 \\ 1 & -1 & 5 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}. \end{aligned}$$

Ejercicios 3.4

1. Para cada una de las matrices dadas a continuación:

$$a.\begin{bmatrix} 1 & 1 & -4 \\ 1 & 3 & -1 \end{bmatrix}. \quad b.\begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}. \quad c.\begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \qquad a, b \in \mathbb{R}.$$

$$d.\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & -1 \end{bmatrix}. \qquad e.\begin{bmatrix} 1 & 3 \\ 3 & 1 \\ -2 & 2 \end{bmatrix}. \qquad f.\begin{bmatrix} 3 & -1 \\ -1 & 1 \\ 4 & -5 \end{bmatrix}.$$

i) Encuentre una descomposición en valores singulares.

ii) Determine la descomposición polar tanto a izquierda como a derecha .

2. Si A es una matriz real de tamaño $n \times n$ con valores singulares todos iguales a 1, muestre que A es ortogonal.

3. Si A es una matriz real de tamaño $n \times n$, ¿cuál es el producto de sus valores singulares $\sigma_1 \cdot \sigma_2 \cdot \ldots \cdot \sigma_n$?

4. Si A es una matriz real de tamaño 2×2 y $\vec{u} \in \mathbb{R}^2$ es unitario, muestre que

$$\sigma_2 \leq \|A\vec{u}\| \leq \sigma_1,$$

donde σ_1, σ_2 son los valores singulares de A.

5. Si A es una matriz real de tamaño $m \times n$ y $\vec{v} \in \mathbb{R}^n$, muestre que

$$\sigma_n\|\vec{v}\| \leq \|A\vec{v}\| \leq \sigma_1\|\vec{v}\|,$$

donde σ_1, σ_n son los valores singulares más grande y más pequeño de la matriz A, respectivamente.

Capítulo 4

Matrices complejas

En la Sección 2.3 se consideraron matrices de componentes reales, en las cuales los valores propios y vectores propios eran complejos. En este capítulo se desarrollará la teoría correspondiente a valores propios y vectores propios para matrices de componentes complejas. El objetivo principal es estudiar algunas factorizaciones para este tipo de matrices, de manera análoga a como vimos en el Capítulo 3.

4.1 Clases especiales de matrices complejas

Los tipos especiales de *matrices cuadradas complejas* que se analizan a continuación son las *hermitianas, antihermitianas* y *unitarias*, por tener características particulares y por ser muy útiles en ingeniería, y en especial en física atómica. Estas matrices generalizan las tres clases de *matrices reales especiales: simétricas, antisimétricas* y *ortogonales*.

4.1.1 Matrices hermitianas

Recordemos que una matriz simétrica $A = [a_{ij}]$ con componentes reales es una matriz que tiene la propiedad de que $A = A^t$. Las *matrices hermitianas* (o hermíticas) son las análogas para el caso en el cual las componentes de la matriz son números complejos.

Definición 4.1 Matriz hermitiana

Se dice que una matriz A de tamaño $n \times n$ es hermitiana si

$$A = A^H. \tag{4.1}$$

Ejemplo 4.1 *Sea A la matriz de componentes complejas*

$$A = \begin{bmatrix} 3 & 4-5i & 3+2i \\ 4+5i & 1 & 7+6i \\ 3-2i & 7-6i & 2 \end{bmatrix}.$$

Comprobar que A es una matriz hermitiana.

Solución.

$$\overline{A} = \begin{bmatrix} \overline{3} & \overline{4-5i} & \overline{3+2i} \\ \overline{4+5i} & \overline{1} & \overline{7+6i} \\ \overline{3-2i} & \overline{7-6i} & \overline{2} \end{bmatrix} = \begin{bmatrix} 3 & 4+5i & 3-2i \\ 4-5i & 1 & 7-6i \\ 3+2i & 7+6i & 2 \end{bmatrix}.$$

$$A^H = \overline{A}^{\,t} = \begin{bmatrix} 3 & 4-5i & 3+2i \\ 4+5i & 1 & 7+6i \\ 3-2i & 7-6i & 2 \end{bmatrix} = A.$$

Nótese que los elementos de la diagonal principal de una matriz hermitiana son números reales, ya que tienen que coincidir con sus conjugados.

Teorema 4.1 *Sea A una matriz hermitiana, entonces para todos los vectores $\vec{x} \in \mathbb{C}^n$, $\vec{x}^H A\vec{x}$ es real.*

Demostración.

La demostración consiste en un cálculo directo

$$\left(\vec{x}^H A\vec{x}\right)^H = \vec{x}^H A^H \left(\vec{x}^H\right)^H = \vec{x}^H A\vec{x},$$

pero como $\vec{x}^H A\vec{x}$ es una matriz hermitiana de tamaño 1×1, se concluye que es un número real. ■

Teorema 4.2 *Si A es una matriz hermitiana, entonces sus valores propios son reales.*

Demostración.

Supongamos que λ es un valor propio y que $\vec{x}$ es un vector propio correspondiente. Es decir,

$$A\vec{x} = \lambda\vec{x}.$$

Si se premultiplica por $\vec{x}^H$, se obtiene

$$\vec{x}^H A\vec{x} = \lambda\vec{x}^H\vec{x}.$$

Pero por el Teorema 4.1, el lado izquierdo es real y la expresión del lado derecho $\vec{x}^H\vec{x} = |\vec{x}|^2 \neq 0$. Se concluye que λ debe ser real. ■

Teorema 4.3 *Sea A una matriz hermitiana de tamaño $n \times n$. Entonces, los vectores propios correspondientes a valores propios distintos de A son ortogonales.*

Demostración.

Sean $\vec{v}_1$ y $\vec{v}_2$ vectores propios asociados a valores propios distintos, digamos, λ_1 y λ_2. Es decir,

$$\begin{aligned} A\vec{v}_1 &= \lambda_1\vec{v}_1 \qquad \text{y} \qquad A\vec{v}_2 = \lambda_2\vec{v}_2 \\ \vec{v}_2^H A\vec{v}_1 &= \lambda_1\vec{v}_2^H\vec{v}_1 \qquad \text{y} \qquad \vec{v}_1^H A\vec{v}_2 = \lambda_2\vec{v}_1^H\vec{v}_2. \end{aligned}$$

Al tomar la transpuesta conjugada de la primera expresión, se tiene

$$\begin{aligned} \big(\vec{v}_2^H A\vec{v}_1\big)^H &= \big(\lambda_1\vec{v}_2^H\vec{v}_1\big)^H \\ \vec{v}_1^H A\vec{v}_2 &= \lambda_1\vec{v}_1^H\vec{v}_2. \end{aligned}$$

En la última expresión se usaron los hechos de que $A^H = A$ y λ_1 es real. Luego, se tiene que

$$\lambda_2\vec{v}_1^H\vec{v}_2 = \lambda_1\vec{v}_1^H\vec{v}_2.$$

Por lo tanto, $\big(\lambda_1 - \lambda_2\big)\vec{v}_1^H\vec{v}_2 = 0$. Pero $\lambda_1 - \lambda_2 \neq 0$, así que $\vec{v}_1^H\vec{v}_2 = 0$. Esto es, $\vec{v}_1$ y $\vec{v}_2$ son ortogonales. ■

Teorema 4.4 *Sea $A = [a_{ij}]$ una matriz con componentes complejas de tamaño $n \times n$, entonces:*

$$\operatorname{tr}\big(AA^H\big) = 0 \qquad \textit{si y solo si} \qquad A = \mathrm{O}\,. \tag{4.2}$$

En realidad, $\operatorname{tr}\big(AA^H\big) > 0$ *si* $A \neq \mathrm{O}\,$.

Demostración.

Si $A = [a_{ij}]$, entonces $A^H = [b_{ij}]$, en donde $b_{ij} = \overline{a}_{ji}$. Si se define $C = AA^H$, se tiene

$$\begin{aligned} c_{ik} &= a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk} \\ &= \sum_{j=1}^{n} a_{ij}b_{jk} = \sum_{j=1}^{n} a_{ij}\overline{a}_{kj}. \end{aligned}$$

En particular, las componentes de la diagonal de C están dadas por

$$c_{ii} = \sum_{j=1}^{n} a_{ij}\overline{a}_{ij} = \sum_{j=1}^{n} \left|a_{ij}\right|^2.$$

Por lo tanto,

$$\operatorname{tr}(C) = \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} \left|a_{ij}\right|^2\right).$$

Como $\left|a_{ij}\right|^2 \geq 0$, la única forma de que esta suma sea cero es que cada $a_{ij} = 0$ para todo i y j. Esto significa que $A = \mathrm{O}$. ■

Teorema 4.5 *Dada una matriz compleja cualquiera A de tamaño $n \times n$, la matriz $H = \frac{1}{2}\left(A + A^H\right)$ es hermitiana.*

Demostración.

Queda como ejercicio para el lector. ■

Definición 4.2 Matriz de proyección espectral compleja

Sea A una matriz compleja de tamaño $n \times n$ que no tiene valores propios múltiples y sean $\vec{v}_k$, $\vec{w}_k$ los vectores propios a derecha e izquierda de A, es decir, los vectores propios de A y A^H asociados a los valores propios λ_k y $\overline{\lambda_k}$, respectivamente; se define la matriz de proyección espectral compleja correspondiente para cada λ_k como:

$$\mathbf{E}(\lambda_k) = \frac{\vec{v}_k\vec{w}_k^H}{\langle \vec{w}_k, \vec{v}_k \rangle} = \frac{\vec{v}_k\vec{w}_k^H}{\vec{w}_k^H\vec{v}_k}. \tag{4.3}$$

Teorema 4.6 *Sea A una matriz compleja de tamaño $n \times n$ que no tiene valores propios múltiples, entonces los vectores propios a derecha e izquierda asociados a valores propios distintos de A son ortogonales.*

Demostración.

Sean $\vec{v}_i$ y $\vec{w}_j$ los vectores propios a derecha e izquierda asociados a valores propios distintos λ_i y λ_j, respectivamente. Para comprobar que $\vec{v}_i \cdot \vec{w}_j = 0$, se calcula

$$\begin{aligned}
\lambda_j(\vec{w}_j \cdot \vec{v}_i) &= (\overline{\lambda_j}\vec{w}_j)^H \vec{v}_i \\
&= (A^H \vec{w}_j)^H \vec{v}_i \quad \text{puesto que } \vec{w}_j \text{ es un vector propio de } A^H \\
&= (\vec{w}_j^H A)\vec{v}_i = \vec{w}_j^H (A\vec{v}_i) \text{ reagrupando términos} \\
&= \vec{w}_j^H (\lambda_i \vec{v}_i) \qquad \text{ya que } \vec{v}_i \text{ es un vector propio de } A \\
&= \lambda_i \vec{w}_j^H \vec{v}_i = \lambda_i (\vec{w}_j \cdot \vec{v}_i).
\end{aligned}$$

Luego, $(\lambda_j - \lambda_i)\vec{w}_j \cdot \vec{v}_i = 0$, y como $\lambda_j - \lambda_i \neq 0$, entonces $\vec{w}_j \cdot \vec{v}_i = 0$. ■

Teorema 4.7 *Sea A una matriz compleja de tamaño $n \times n$ con valores propios distintos $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces las matrices de proyección espectral $\mathbf{E}(\lambda_k)$ definidas en (4.3) satisfacen las siguientes propiedades:*

$$a)\ \mathbf{E}(\lambda_i)\mathbf{E}(\lambda_j) = \begin{cases} \mathbf{E}(\lambda_i) & \text{si } i = j, \\ \mathrm{O} & \text{si } i \neq j. \end{cases} \qquad b)\ \sum_{k=1}^{n} \mathbf{E}(\lambda_k) = I_n,$$

c) Cada $\mathbf{E}(\lambda_k)$ conmuta con A, es decir, $A\mathbf{E}(\lambda_k) = \mathbf{E}(\lambda_k)A$.

Demostración.

a) Sean $\vec{v}_i$ y $\vec{w}_j$ los vectores propios a derecha e izquierda asociados a valores propios distintos λ_i y λ_j, respectivamente; por el Teorema 4.6 estos vectores son ortogonales. Por lo tanto, de la definición de $\mathbf{E}(\lambda_k)$ dada en (4.3), se tiene

$$\mathbf{E}(\lambda_i)\mathbf{E}(\lambda_j) = \frac{\vec{v}_i \vec{w}_i^H}{\vec{w}_i^H \vec{v}_i} \frac{\vec{v}_j \vec{w}_j^H}{\vec{w}_j^H \vec{v}_j} = \frac{\vec{v}_i}{\vec{w}_i^H \vec{v}_i} \vec{w}_i^H \vec{v}_j \frac{\vec{w}_j^H}{\vec{w}_j^H \vec{v}_j} = \mathrm{O},$$

y de manera análoga

$$\mathbf{E}(\lambda_i)\mathbf{E}(\lambda_i) = \frac{\vec{v}_i\vec{w}_i^H}{\vec{w}_i^H\vec{v}_i}\,\frac{\vec{v}_i\vec{w}_i^H}{\vec{w}_i^H\vec{v}_i} = \vec{v}_i\frac{\vec{w}_i^H\vec{v}_i}{\vec{w}_i^H\vec{v}_i}\frac{\vec{w}_i^H}{\vec{w}_i^H\vec{v}_i} = \mathbf{E}(\lambda_i).$$

b) Queda como ejercicio para el lector.

c) Al premultiplicar por A cualquier matriz $\mathbf{E}(\lambda_k)$, se obtiene

$$A\mathbf{E}(\lambda_k) = A\vec{v}_k\frac{\vec{w}_k^H}{\vec{v}_k^H\vec{w}_k} = \lambda_k\frac{\vec{v}_k\vec{w}_k^H}{\vec{v}_k^H\vec{w}_k} = \lambda_k\mathbf{E}(\lambda_k), \tag{4.4}$$

y al multiplicar por A cualquier matriz $\mathbf{E}(\lambda_k)$, se llega a

$$\mathbf{E}(\lambda_k)A = \frac{\vec{v}_k}{\vec{v}_k^H\vec{w}_k}\big(A^H\vec{w}_k\big)^H = \frac{\vec{v}_k}{\vec{v}_k^H\vec{w}_k}\big(\overline{\lambda_k}\vec{w}_k\big)^H = \lambda_k\mathbf{E}(\lambda_k). \tag{4.5}$$

Luego, $A\mathbf{E}(\lambda_k) = \mathbf{E}(\lambda_k)A$ para $k = 1, 2, \ldots, n$. ■

Ejemplo 4.2 *Encuentre las proyecciones espectrales de la matriz:*

$$A = \begin{bmatrix} 1-3i & 2 & -2-2i \\ 1-i & 2 & 1+i \\ -2+i & -1-2i & 2+3i \end{bmatrix}.$$

Solución.

La ecuación característica asociada a la matriz A es

$$p_A(\lambda) = -\lambda^3 + 5\lambda^2 - 8\lambda + 6 = -(\lambda - 3)(\lambda^2 - 2\lambda + 2) = 0,$$

entonces se tiene que los valores propios asociados a la matriz A son $\lambda_1 = 3,\ \lambda_2 = 1+i$ y $\lambda_3 = 1-i$, y los respectivos vectores propios de A son

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 1+i \\ 1 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} -1 \\ 1-i \\ 1 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}.$$

Los vectores de A^H asociados a $\lambda_1,\ \overline{\lambda_2},\ \overline{\lambda_3}$ son

$$\vec{w}_1 = \begin{bmatrix} i \\ 1 \\ -1 \end{bmatrix}, \qquad \vec{w}_2 = \begin{bmatrix} 1 \\ -i \\ 1+i \end{bmatrix} \qquad \text{y} \qquad \vec{w}_3 = \begin{bmatrix} 2 \\ -i \\ 1+i \end{bmatrix}.$$

Luego, las matrices de proyección espectral $\mathbf{E}(\lambda_k)$ son

$$\mathbf{E}(\lambda_1) = \frac{1}{i}\begin{bmatrix} 0 \\ 1+i \\ 1 \end{bmatrix}\begin{bmatrix} -i & 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -1-i & 1-i & -1+i \\ -1 & -i & i \end{bmatrix},$$

$$\mathbf{E}(\lambda_2) = \begin{bmatrix} -1 \\ 1-i \\ 1 \end{bmatrix}\begin{bmatrix} 1 & i & 1-i \end{bmatrix} = \begin{bmatrix} -1 & -i & -1+i \\ 1-i & 1+i & -2i \\ 1 & i & 1-i \end{bmatrix}$$

$$\mathbf{E}(\lambda_3) = \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}\begin{bmatrix} 2 & i & 1-i \end{bmatrix} = \begin{bmatrix} 2 & i & 1-i \\ 2i & -1 & 1+i \\ 0 & 0 & 0 \end{bmatrix}.$$

Nótese que $\mathbf{E}(\lambda_1) + \mathbf{E}(\lambda_2) + \mathbf{E}(\lambda_3) = I_3$.

4.1.2 Matrices antihermitianas

Como se ha visto, una matriz antisimétrica A es una matriz real que tiene la propiedad de que $A^t = -A$. Las matrices antihermitianas constituyen el análogo para el caso complejo.

Definición 4.3 Matriz antihermitiana

Se dice que una matriz A de tamaño $n \times n$ es antihermitiana si:

$$A^H = -A. \tag{4.6}$$

Los elementos de la diagonal principal de una matriz antihermitiana son cero o imaginarios puros, ya que tienen que coincidir con sus conjugados.

Teorema 4.8 *Si A es antihermitiana, entonces para todos los vectores complejos $\vec{z}$, $\vec{z}^H A\vec{z}$, es cero o imaginario puro*[1].

Demostración.

La demostración consiste en un cálculo directo

$$(\vec{z}^H A\vec{z})^H = \vec{z}^H A^H (\vec{z}^H)^H = -\vec{z}^H A\vec{z}.$$

[1] Un imaginario puro es un número complejo de la forma αi con α real.

Si se expresa $\vec{z}^H A\vec{z} = \alpha + i\beta$, entonces la ecuación anterior se puede escribir como

$$\alpha - i\beta = -\big(\alpha + i\beta\big),$$

luego se debe tener que $\alpha = -\alpha$, así que $\alpha = 0$. Por lo tanto, se concluye que $\vec{z}^H A\vec{z}$ es un imaginario puro. ■

Teorema 4.9 *Los valores propios de una matriz antihermitiana deben ser cero o imaginarios puros.*

Demostración.

Supongamos que λ es un valor propio y que $\vec{z}$ es el vector propio correspondiente. Es decir,

$$A\vec{z} = \lambda\vec{z}.$$

Si se premultiplica por $\vec{z}^H$, se obtiene

$$\vec{z}^H A\vec{z} = \lambda\vec{z}^H\vec{z}.$$

Pero por el Teorema 4.8, el lado izquierdo es cero o imaginario puro y el lado derecho $\vec{z}^H\vec{z} = |\vec{z}|^2$ es real y distinto de cero. Por lo tanto,

$$\lambda = \frac{\vec{z}^H A\vec{z}}{\vec{z}^H\vec{z}},$$

luego λ debe ser cero o imaginario puro. ■

Teorema 4.10 *Sea A una matriz antihermitiana de tamaño $n \times n$. Entonces los vectores propios asociados con valores propios distintos de A son ortogonales.*

Demostración.

Sean $\vec{v}_1$ y $\vec{v}_2$ vectores propios de A que corresponden a valores propios distintos, digamos, λ_1 y λ_2. Para probar que $\vec{v}_1 \cdot \vec{v}_2 = 0$, se calcula

$$\begin{aligned}
\overline{\lambda_1}\vec{v}_1 \cdot \vec{v}_2 &= \big(\lambda_1\vec{v}_1\big)^H\vec{v}_2 = \big(A\vec{v}_1\big)^H\vec{v}_2 && \text{puesto que } \vec{v}_1 \text{ es un vector propio}\\
&= \big(\vec{v}_1^H A^H\big)\vec{v}_2 = -\vec{v}_1^H(A\vec{v}_2) && \text{puesto que } A \text{ es antihermitiana}\\
&= -\,\vec{v}_1^H(\lambda_2\vec{v}_2) && \text{puesto que } \vec{v}_2 \text{ es un vector propio}\\
&= -\,\lambda_2\vec{v}_1^H\vec{v}_2 = -\lambda_2\vec{v}_1 \cdot \vec{v}_2.
\end{aligned}$$

Por lo tanto, se tiene que $\big(\overline{\lambda_1} + \lambda_2\big)\vec{v}_1 \cdot \vec{v}_2 = 0$. Pero como $\overline{\lambda_1} = -\lambda_1$ por ser imaginario puro, entonces $\overline{\lambda_1} + \lambda_2 \neq 0$, así que $\vec{v}_1 \cdot \vec{v}_2 = 0$. ■

Teorema 4.11 *Si A es hermitiana, entonces $B = iA$ es antihermitiana. Análogamente, si B es una matriz antihermitiana, entonces la matriz $A = iB$ es hermitiana.*

Demostración.
Sea $A^H = A$ y definamos $B = iA$, entonces

$$B^H = (iA)^H = (i)^H\ A^H = (-i)(A) = -B.$$

Esto prueba que B es antihermitiana. De la misma manera, se puede probar que $A = iB$ es hermitiana cuando B es antihermitiana. ■

Teorema 4.12 *Dada una matriz compleja cualquiera A de tamaño $n \times n$, la matriz $S = \frac{1}{2}\big(A - A^H\big)$ es antihermitiana.*

Demostración.
Queda como ejercicio para el lector. ■

4.1.3 Matrices unitarias

Recordemos que una matriz ortogonal A es una matriz real que tiene la propiedad de que $A^t = A^{-1}$. Las matrices unitarias son el análogo para el caso complejo.

Definición 4.4 Matriz unitaria

Una matriz cuadrada U de componentes complejas se dice que es una matriz unitaria si $U^H U = I$. En consecuencia, U es no singular y se tiene $U^{-1} = U^H$.

Teorema 4.13 *Sea U una matriz unitaria, entonces sus valores propios son de módulo igual a* 1.

Demostración.

Sea λ un valor propio de U con vector propio asociado $\vec{v}$, es decir

$$U\vec{v} = \lambda\vec{v}. \tag{4.7}$$

Luego, tomando la transpuesta conjugada, se tiene

$$\begin{aligned} (U\,\vec{v})^H &= \big(\lambda\,\vec{v}\big)^H \\ \vec{v}^H\,U^H &= \overline{\lambda}\,\vec{v}^H, \end{aligned}$$

pero como U es unitaria, se tiene que

$$\vec{v}^H\,U^{-1} = \overline{\lambda}\,\vec{v}^H. \tag{4.8}$$

Si se multiplica por la derecha ambos lados de (4.8) por $U\vec{v}$, se obtiene

$$\begin{aligned} \vec{v}^H\,U^{-1}(U\vec{v}) &= \overline{\lambda}\,\vec{v}^H(U\vec{v}) \\ \vec{v}^H(U^{-1}U)\vec{v} &= \overline{\lambda}\,\vec{v}^H(\lambda\vec{v}) && \text{por (4.7)} \\ \vec{v}^H\vec{v} &= \overline{\lambda}\,\lambda\,\vec{v}^H\vec{v}. \end{aligned}$$

Pero como $\vec{v}^H\vec{v} \neq 0$, se concluye que $\overline{\lambda}\,\lambda = 1$. Es decir, $\|\lambda\|^2 = 1$, así que $\|\lambda\| = 1$. ■

Teorema 4.14 *Si U es una matriz unitaria de tamaño $n \times n$. Entonces los vectores propios asociados con valores propios distintos de U son ortogonales.*

Demostración.

Sean $\vec{v}_1$ y $\vec{v}_2$ vectores propios de U que corresponden a valores propios distintos, digamos, λ_1 y λ_2. Para demostrar que $\vec{v}_1 \cdot \vec{v}_2 = 0$, se calcula

$$\begin{aligned} \big(U\,\vec{v}_1\big)^H\big(U\,\vec{v}_2\big) &= \big(\lambda_1\,\vec{v}_1\big)^H\big(\lambda_2\,\vec{v}_2\big) && \text{puesto que } \lambda_1, \lambda_2 \text{ son valores propios} \\ \big(\vec{v}_1^H\,U^H\big)\big(U\vec{v}_2\big) &= \overline{\lambda}_1\,\vec{v}_1^H\lambda_2\,\vec{v}_2 && \text{puesto que } U \text{ es unitaria} \\ \vec{v}_1^H\,\vec{v}_2 &= \overline{\lambda}_1\,\lambda_2\,\vec{v}_1^H\,\vec{v}_2. \end{aligned}$$

Por lo tanto, $\big(1 - \overline{\lambda}_1\,\lambda_2\big)\vec{v}_1 \cdot \vec{v}_2 = 0$. Pero como λ_1 es distinto de λ_2, entonces $\overline{\lambda}_1\,\lambda_2 \neq 1$, así que $\vec{v}_1 \cdot \vec{v}_2 = 0$. ■

4.1.4 Matrices normales

Definición 4.5 Matriz normal

Se dice que la matriz de componentes complejas N de tamaño $n \times n$ es normal si conmuta con N^H, es decir:

$$NN^H = N^H N.$$

Ejemplo 4.3 *Comprobar que las matrices complejas diagonales son normales.*

Solución.
Sea D la siguiente matriz diagonal de tamaño $n \times n$

$$D = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\},$$

entonces

$$\begin{aligned} DD^H &= diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\} diag\{\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_n\} \\ &= diag\{|\lambda_1|^2, |\lambda_2|^2, \ldots, |\lambda_n|^2\} \\ &= diag\{\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_n\} diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\} \\ &= D^H D, \end{aligned}$$

por lo tanto D es una matriz normal.

Teorema 4.15 *Las matrices hermitianas, las antihermitianas y las unitarias son matrices normales.*

Demostración.
Supongamos que A es hermitiana, entonces

$$A^H A = AA = A^2 \qquad \text{y} \qquad AA^H = AA = A^2.$$

Luego, $AA^H = A^H A$.
Las demás quedan como ejercicio para el lector. ■

Teorema 4.16 *Dadas A y B matrices complejas de tamaño $n \times n$, si $AB = BA$ entonces A y B tienen un vector propio en común.*

Demostración.

La prueba es análoga a la del Teorema 3.8 y queda como ejercicio para el lector. ■

4.2 Factorizaciones

En esta sección se explica cómo se puede expresar una matriz A de componentes complejas como el producto de dos o más matrices.

Definición 4.6 Matrices complejas semejantes

Una matriz de componentes complejas A de tamaño $n \times n$ es semejante a una matriz de componentes complejas B de tamaño $n \times n$ si existe una matriz de componentes complejas no singular P de tamaño $n \times n$ tal que

$$B = P^{-1}AP. \tag{4.9}$$

De manera análoga, se dice que A y B son semejantes si y solo si existe una matriz de componentes complejas no singular P tal que

$$PB = AP. \tag{4.10}$$

Teorema 4.17 *Las matrices complejas semejantes tienen el mismo polinomio característico y, por tanto, los mismos valores propios.*

Demostración.

Como A y B son matrices complejas semejantes de tamaño $n \times n$, $B = P^{-1}AP$, entonces

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}\big[AP - \lambda P\big] = P^{-1}\big[A - \lambda I\big]P.$$

Por consiguiente,

$$\begin{aligned} \det\big(B - \lambda I\big) &= \det\big[P^{-1}(A - \lambda I)P\big] = \det\big(P^{-1}\big)\det(A - \lambda I)\det(P) \\ &= \det\big(P^{-1}\big)\det(P)\det(A - \lambda I) = \det(A - \lambda I). \end{aligned}$$

Esto significa que A y B tienen la misma ecuación característica, y como los valores propios son raíces de la ecuación característica, tienen los mismos valores propios. ■

Definición 4.7 Matrices congruentes hermitianas

Dos matrices hermitianas A y B de tamaño $n \times n$ son congruentes hermitianas si existe una matriz P no singular de componentes complejas de tamaño $n \times n$ tal que

$$A = P^H B P. \tag{4.11}$$

Teorema 4.18 Descomposición de Sylvester

Sea A una matriz compleja de tamaño $n \times n$ con valores propios distintos $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces A se puede escribir como

$$A = \sum_{k=1}^{n} \lambda_k \mathbf{E}(\lambda_k), \tag{4.12}$$

donde la matriz $\mathbf{E}(\lambda_k)$ es dada en (4.3).

Demostración.

Puesto que cada $\mathbf{E}(\lambda_k)$ conmuta con A, al sumar las expresiones obtenidas en (4.4) y (4.5), se tiene

$$\underbrace{\sum_{k=1}^{n} A\mathbf{E}(\lambda_k)}_{} = \underbrace{\sum_{k=1}^{n} \mathbf{E}(\lambda_k)A}_{} = \sum_{k=1}^{n} \lambda_k \mathbf{E}(\lambda_k)$$

$$A \sum_{k=1}^{n} \mathbf{E}(\lambda_k) = \left[\sum_{k=1}^{n} \mathbf{E}(\lambda_k)\right] A = \sum_{k=1}^{n} \lambda_k \mathbf{E}(\lambda_k),$$

y usando la propiedad b) del Teorema 4.7, se completa la prueba. ■

Ejemplo 4.4 *Encuentre la descomposición de Sylvester para la matriz dada en el Ejemplo 4.2.*

Solución.

En el Ejemplo 4.2, se obtuvo que los valores propios asociados a la matriz A eran $\lambda_1 = 3,\ \lambda_2 = 1+i,\ \lambda_3 = 1-i$, y las matrices complejas de proyección espectral $\mathbf{E}\big(\lambda_k\big)$ fueron

$$\mathbf{E}\big(\lambda_1\big) = \begin{bmatrix} 0 & 0 & 0 \\ -1-i & 1-i & -1+i \\ -1 & -i & i \end{bmatrix}, \quad \mathbf{E}\big(\lambda_2\big) = \begin{bmatrix} -1 & -i & -1+i \\ 1-i & 1+i & -2i \\ 1 & i & 1-i \end{bmatrix},$$

$$\mathbf{E}\big(\lambda_3\big) = \begin{bmatrix} 2 & i & 1-i \\ 2i & -1 & 1+i \\ 0 & 0 & 0 \end{bmatrix},$$

luego

$$\sum_{i=1}^{3} \lambda_i \mathbf{E}\big(\lambda_i\big) = 3\mathbf{E}\big(\lambda_3\big) + \big(1+i\big)\mathbf{E}\big(\lambda_3\big)\big(1-i\big)\mathbf{E}\big(\lambda_3\big),$$

la cual coincide con la matriz A.

Teorema 4.19 Teorema de Schur

Sea A una matriz compleja de tamaño $n \times n$. Entonces A es semejante a una matriz triangular superior T, mediante una matriz unitaria U, es decir

$$T = U^H A U.$$

Entonces, se dice que A es triangularizable por una matriz unitaria U.

Demostración.

La demostración es por inducción sobre n. Si $n = 1$, A es una matriz de tamaño 1×1 que es triangular. La matriz unitaria es $U = [1]$.

Supongamos que toda matriz de componentes complejas de tamaño $\big(n-1\big) \times \big(n-1\big)$ es triangularizable por una matriz unitaria. Sea A una matriz de tamaño $n \times n$. Sabemos que su polinomio característico tiene al menos una raíz compleja λ_1. Sea $\vec{v}_1 \in \mathbb{C}^n$ un vector propio normalizado asociado al valor propio λ_1. Denotemos por $\mathbb{W}$ el complemento ortogonal

a $\vec{v}_1$ de dimensión $n-1$. Sea $\{\vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n\}$ una base ortonormal de $\mathbb{W}$. Luego, cada vector $\vec{X}$ de $\mathbb{W}$ tiene la forma

$$\vec{X} = a_2\ \vec{v}_2 + a_3\ \vec{v}_3 + \ldots + a_n\ \vec{v}_n.$$

La matriz de cambio de base, de la *base canónica* de $\mathbb{C}^n$ a la base $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, es la matriz S cuyas columnas son los elementos de los vectores $\vec{v}_i$. Luego,

$$\begin{aligned} AS &= \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}. \end{aligned}$$

Por lo tanto,

$$S^{-1}AS = S^{-1}\begin{bmatrix} \lambda_1\vec{v}_1 & A\vec{v}_2 & \ldots & A\vec{v}_n \end{bmatrix}.$$

Pero como S es unitaria, se tiene que $S^{-1} = S^H$. Por consiguiente

$$S^H AS = \begin{bmatrix} \lambda_1 & \overline{z_{12}} & \ldots & \overline{z_{1n}} \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{bmatrix},$$

donde $z_{1i} = \vec{v}_1^H A\vec{v}_i$ y A_1 es una matriz de tamaño $\big(n-1\big) \times \big(n-1\big)$.

La prueba se completa por inducción, sea R_1 una matriz unitaria de tamaño $(n-1) \times (n-1)$ tal que $(R_1)^H A_1 R_1 = T_1$, con T_1 triangular superior por la hipótesis de inducción. Entonces, la matriz

$$R = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & R_1 & \\ 0 & & & \end{bmatrix}$$

es unitaria y

$$\begin{aligned} \big(SR\big)^H A\big(SR\big) &= R^H\big(S^H AS\big)R \\ &= \begin{bmatrix} 1 & \vec{0}^t \\ \vec{0} & R_1^H \end{bmatrix} \begin{bmatrix} \lambda_1 & \vec{z}^H \\ \vec{0} & A_1^H \end{bmatrix} \begin{bmatrix} 1 & \vec{0}^t \\ \vec{0} & R_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \vec{z}^H R_1 \\ \vec{0} & T_1 \end{bmatrix}, \end{aligned}$$

donde $\vec{z}^H = [\overline{z_{12}} \ \ \overline{z_{13}} \ \ \ldots \ \ \overline{z_{1n}}]$. La matriz $SR = U$ es el producto de dos matrices unitarias; por lo tanto, es también una matriz unitaria. Así, $U^H AU$ es una matriz triangular superior y nuestra prueba queda completa. ■

El siguiente resultado corresponde a una consecuencia directa del teorema anterior.

Teorema 4.20 *Sea A una matriz de componentes complejas de tamaño $n \times n$. Los valores propios de A son los elementos de la diagonal de la matriz triangular superior T semejante a A por una matriz unitaria.*

Demostración.

Como A y T son semejantes por el Teorema 4.17, tienen el mismo polinomio característico. Por otra parte, como T es triangular, se tiene

$$p_A(\lambda) = p_T(\lambda) = \big(t_{11} - \lambda\big)\big(t_{22} - \lambda\big) \ldots \big(t_{nn} - \lambda\big),$$

donde $t_{11}, t_{22}, \ldots, t_{nn}$ son los elementos de la diagonal de T. Así pues los valores propios de A son los elementos de la diagonal de T. ■

Ejemplo 4.5 *Dada la matriz de componentes complejas:*

$$A = \begin{bmatrix} 1-i & 0 & 0 \\ i & 1 & 2+i \\ 2i & 0 & i \end{bmatrix},$$

encuentre una matriz T que sea la triangularización de A.

Solución.

El polinomio característico de A es

$$p_A(\lambda) = -\lambda^3 + 2\lambda^2 - (2+i)\lambda + (1+i) = -(1-\lambda)(i-\lambda)((1-i)-\lambda).$$

Luego, los valores propios son $\lambda_1 = 1$, $\lambda_2 = i$ y $\lambda_3 = 1 - i$. El vector propio correspondiente a $\lambda_1 = 1$ es $\vec{v}_1 = \begin{bmatrix} 0 \\ 3+4i \\ 0 \end{bmatrix}$; para $\lambda_2 = i$ es $\vec{v}_2 = \begin{bmatrix} 0 \\ -1-3i \\ 2 \end{bmatrix}$ y, por último, para $\lambda_3 = 1 - i$ es $\vec{v}_3 = \begin{bmatrix} 1-2i \\ -5 \\ 2i \end{bmatrix}$.

Para determinar U, se aplica el proceso de Gram-Schmidt a $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ para encontrar una base ortonormal para $\mathbb{C}^3$, puesto que el producto punto en complejos no es conmutativo, entonces hay que tener en cuenta el orden en que se realiza el producto punto. Luego, para determinar los vectores ortogonales, se usa la siguiente expresión:

$$\vec{w}_k = \vec{v}_k - \sum_{i=1}^{k-1} \frac{\langle \vec{w}_i, \vec{v}_k \rangle}{\|\vec{w}_i\|^2} \vec{w}_i = \vec{v}_k - \sum_{i=1}^{k-1} \frac{\vec{w}_i^H \vec{v}_k}{\vec{w}_i^H \vec{w}_i} \vec{w}_i,$$

y la base ortonormal se obtiene dividiendo cada $\vec{w}_k$ por su norma. Como $\|\vec{v}_1\| = 5$, se hace $\vec{u}_1 = \dfrac{\vec{v}_1}{\|\vec{v}_1\|} = \dfrac{1}{5}\begin{bmatrix} 0 \\ 3+4i \\ 0 \end{bmatrix}$, los otros vectores serían

$$\begin{aligned} \vec{w}_2 &= \vec{v}_2 - \big(\vec{u}_1^H \vec{v}_2\big)\vec{u}_1 = \begin{bmatrix} 0 \\ -1-3i \\ 2 \end{bmatrix} - \frac{-15-5i}{25}\begin{bmatrix} 0 \\ 3+4i \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -1-3i \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1+3i \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}. \end{aligned}$$

Entonces, $\|\vec{w}_2\| = 2$ y $\vec{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Se puede verificar que $\vec{u}_1^H \vec{u}_2 = 0$. Ahora,

$$\begin{aligned} \vec{w}_3 &= \vec{v}_3 - \vec{u}_1^H \vec{v}_3 \vec{u}_1 - \vec{u}_2^H \vec{v}_3 \vec{u}_2 \\ &= \begin{bmatrix} 1-2i \\ -5 \\ 2i \end{bmatrix} - \frac{-15+20i}{25}\begin{bmatrix} 0 \\ 3+4i \\ 0 \end{bmatrix} - 2i\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \\ &= \begin{bmatrix} 1-2i \\ -5 \\ 2i \end{bmatrix} + \begin{bmatrix} 0 \\ 5 \\ -2i \end{bmatrix} = \begin{bmatrix} 1-2i \\ 0 \\ 0 \end{bmatrix}. \end{aligned}$$

Por último, $\|\vec{v}_3\| = \sqrt{5}$, luego $\vec{u}_3 = \dfrac{1}{\sqrt{5}}\begin{bmatrix} 1-2i \\ 0 \\ 0 \end{bmatrix}$. También se verifica que la base obtenida para $\mathbb{C}^3$ es ortonormal observando que $\vec{u}_1^H \vec{u}_3 = 0$ y $\vec{u}_2^H \vec{u}_3 = 0$. Por lo tanto, la matriz unitaria U es

$$U = \begin{bmatrix} 0 & 0 & \big(1-2i\big)/\sqrt{5} \\ \big(3+4i\big)/5 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

y la matriz triangular es

$$U^H AU = \begin{bmatrix} 0 & \dfrac{3-4i}{5} & 0 \\ 0 & 0 & 1 \\ \dfrac{1+2i}{\sqrt{5}} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1-i & 0 & 0 \\ i & 1 & 2+i \\ 2i & 0 & i \end{bmatrix} \begin{bmatrix} 0 & 0 & \dfrac{1-2i}{\sqrt{5}} \\ \dfrac{3+4i}{5} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Al realizar los productos, se obtiene

$$T = \begin{bmatrix} 1 & 2-i & \dfrac{2-i}{\sqrt{5}} \\ 0 & i & \dfrac{4+2i}{\sqrt{5}} \\ 0 & 0 & 1-i \end{bmatrix}.$$

Nótese que los elementos de la diagonal principal de la matriz T son los valores propios de la matriz A.

Teorema 4.21 *Sea A una matriz hermítica de tamaño $n \times n$, entonces existe una matriz unitaria U tal que*

$$U^{-1}AU$$

es una matriz diagonal.

Demostración.

Como A es una matriz compleja, por el teorema de Schur, A se puede triangularizar mediante una matriz unitaria U, es decir

$$T = U^H AU,$$

donde T es una matriz triangular superior.

Al tomar la transpuesta conjugada y usando que $A^H = A$, se tiene que

$$T^H = \left(U^H AU\right)^H = U^H A^H \left(U^H\right)^H = U^H AU = T,$$

como T^H es una matriz triangular inferior, luego T es una matriz diagonal. En consecuencia, A es semejante mediante una matriz unitaria U, a una matriz diagonal T. ■

Teorema 4.22 *Si A y B son matrices complejas de tamaño $n \times n$ y $AB = BA$, entonces A y B pueden ser triangularizadas unitariamente en forma simultánea.*

Demostración.

La demostración es por inducción sobre n. Por el Teorema 4.16, existe $\vec{v}_1 \in \mathbb{C}^n$ normalizado tal que

$$A\vec{v}_1 = \lambda_1 \vec{v}_1 \qquad \text{y} \qquad B\vec{v}_1 = \mu_1 \vec{v}_1.$$

De manera análoga, como en la prueba del Teorema 4.19, se construye una matriz S unitaria cuya primera columna sea $\vec{v}_1$. Entonces

$$S^H AS = \begin{bmatrix} \lambda_1 & \overline{x_{12}} & \cdots & \overline{x_{1n}} \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{bmatrix} \quad \text{y} \quad S^H BS = \begin{bmatrix} \mu_1 & \overline{y_{12}} & \cdots & \overline{y_{1n}} \\ 0 & & & \\ \vdots & & B_1 & \\ 0 & & & \end{bmatrix},$$

donde A_1 y B_1 son matrices complejas de tamaño $\big(n-1\big) \times \big(n-1\big)$. Por otra parte,

$$\begin{aligned} \big(S^H AS\big)\big(S^H BS\big) &= S^H (AB) S && \text{puesto que } S^H = S^{-1}, \\ &= S^H \big(BA\big) S && \text{puesto que } AB = BA, \\ &= \big(S^H BS\big)\big(S^H AS\big), \end{aligned}$$

y por la multiplicación por bloques, se tiene

$$\begin{bmatrix} \lambda_1\mu_1 & \vdots & \lambda_1 \vec{y}^H + \vec{x}^H B_1 \\ \cdots & \cdot & \cdots\cdots \\ \vec{0} & \vdots & A_1 B_1 \end{bmatrix} = \begin{bmatrix} \lambda_1\mu_1 & \vdots & \mu_1 \vec{x}^H + \vec{y}^H A_1 \\ \cdots & \cdot & \cdots\cdots \\ \vec{0} & \vdots & B_1 A_1 \end{bmatrix},$$

donde $\vec{x}^H = [\overline{x_{12}} \;\; \overline{x_{13}} \;\; \cdots \;\; \overline{x_{1n}}]$, $\vec{y}^H = [\overline{y_{12}} \;\; \overline{y_{13}} \;\; \cdots \;\; \overline{y_{1n}}]$ y, por lo tanto, $A_1 B_1 = B_1 A_1$. Los detalles para culminar la prueba son análogos a los del Teorema 4.19 y se dejan como ejercicio para el lector. ■

Teorema 4.23 *Si A y B son matrices normales de tamaños $n \times n$, entonces $AB = BA$ si y solo si existe una matriz unitaria U de tamaño $n \times n$ tal que $U^H AU$ y $U^H BU$ sean ambas diagonales.*

Demostración.

Supongamos que $AB = BA$. Por el Teorema 4.22, existe una matriz unitaria U tal que U^HAU y U^HBU son ambas triangulares. Pero como A y B son normales, pueden ser diagonalizables unitariamente. De aquí que U^HAU y U^HBU sean diagonales. Por otra parte, si $U^HAU = D_A$ y $U^HBU = D_B$ son ambas diagonales, entonces

$$\begin{aligned} AB &= \big(UD_AU^H\big)\big(UD_BU^H\big) = UD_AD_BU^H \\ &= UD_BD_AU^H = \big(UD_BU^H\big)\big(UD_AU^H\big) \\ &= BA. \end{aligned}$$

Esto completa la demostración. ■

Ejemplo 4.6 *Considere la matriz de componentes complejas:*

$$\begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix}.$$

Comprobar que es diagonalizable mediante una matriz unitaria.

Solución.

El polinomio característico de A es $p_A(\lambda) = \lambda(3-\lambda)(2+\lambda)$. En este caso, los valores propios de A son $\lambda_1 = 3$, $\lambda_2 = -2$ y $\lambda_3 = 0$. Los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 1 \\ i \\ -1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 2 \\ -i \\ 1 \end{bmatrix}$, respectivamente.

Para encontrar la matriz unitaria U, se ortonormaliza el conjunto $\big\{\vec{v}_1, \vec{v}_2, \vec{v}_3\big\}$. Como $\|\vec{v}_1\| = \sqrt{3}$, se hace $\vec{u}_1 = \dfrac{\vec{v}_1}{\|\vec{v}_1\|} = \dfrac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ i \\ -1 \end{bmatrix}$. Por otra parte, se tiene que $\|\vec{v}_2\| = \sqrt{2}$, entonces $\vec{u}_2 = \dfrac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix}$. Por último, $\|\vec{v}_3\| = \sqrt{6}$ de manera que $\vec{u}_3 = \dfrac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ -i \\ 1 \end{bmatrix}$. Se puede verificar que la base

obtenida para $\mathbb{C}^3$ es ortonormal observando que $\vec{u}_1^H\vec{u}_2 = 0$, $\vec{u}_1^H\vec{u}_3 = 0$ y $\vec{u}_2^H\vec{u}_3 = 0$. Por lo tanto,

$$U = \begin{bmatrix} 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ i/\sqrt{3} & 1/\sqrt{2} & -i/\sqrt{6} \\ -1/\sqrt{3} & -i/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}.$$

Como el $\det(U) = 1$, se tiene que

$$D = U^H AU = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Luego, la matriz A es diagonalizable por una matriz unitaria.

Teorema 4.24 Descomposición espectral para hermitianas

Sea A una matriz hermitiana de tamaño $n \times n$ con valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces A se puede escribir como

$$A = \sum_{i=1}^{n} \lambda_i \mathbf{E}(\lambda_i), \tag{4.13}$$

donde las matrices de proyección espectral vienen dadas por

$$\mathbf{E}(\lambda_i) = \frac{\vec{v}_i\vec{v}_i^H}{\vec{v}_i^H\vec{v}_i} = \vec{u}_i\vec{u}_i^H, \tag{4.14}$$

con $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ los vectores propios normalizados de A.

Demostración.

Por el Teorema 4.21, existe una matriz U tal que $U^{-1}AU = T$, donde T es una matriz diagonal. Entonces,

$$\begin{aligned} A &= UTU^{-1} = UTU^H \\ &= \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \ldots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^H \\ \vec{u}_2^H \\ \vdots \\ \vec{u}_n^H \end{bmatrix}, \end{aligned}$$

luego

$$A = UTU^{-1} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \dots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 \vec{u}_1^H \\ \lambda_2 \vec{u}_2^H \\ \vdots \\ \lambda_n \vec{u}_n^H \end{bmatrix}$$
$$= \lambda_1 \vec{u}_1 \vec{u}_1^H + \lambda_2 \vec{u}_2 \vec{u}_2^H + \dots + \lambda_n \vec{u}_n \vec{u}_n^H.$$

Esto prueba el teorema. ■

Ejemplo 4.7 *Ilustrar el teorema de descomposición espectral para la matriz dada en el Ejemplo 4.6.*

Solución.

Del Ejemplo 4.6 se tiene que los valores propios asociados a la matriz A son $\lambda_1 = 3$, $\lambda_2 = -2$ y $\lambda_3 = 0$. Los respectivos vectores propios normalizados de A eran

$$\vec{u}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ i \\ -1 \end{bmatrix}, \qquad \vec{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix} \qquad \text{y} \qquad \vec{u}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ -i \\ 1 \end{bmatrix}.$$

Entonces

$$\sum_{i=1}^{3} \lambda_i \vec{u}_i \vec{u}_i^H = \frac{3}{3} \begin{bmatrix} 1 \\ i \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -i & -1 \end{bmatrix} + \frac{-2}{2} \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix} \begin{bmatrix} 0 & 1 & i \end{bmatrix}$$
$$= \begin{bmatrix} 1 & -i & -1 \\ i & 1 & -i \\ -1 & i & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & i \\ 0 & -i & 1 \end{bmatrix} = \begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix},$$

la cual coincide con la matriz A dada en el Ejemplo 4.6.

Teorema 4.25 *Sea A una matriz hermitiana y no singular que se puede factorizar como*

$$A = LDU$$

sin intercambio de filas, donde L es una matriz unitaria triangular inferior, D es una matriz diagonal y U es una matriz unitaria triangular superior. Entonces, $L = U^H$.

Demostración.

Puesto que A puede factorizarse como $A = LDU$, tomando la transpuesta conjugada, se tiene que

$$A^H = \big(LDU\big)^H = U^H D^H L^H = U^H D L^H.$$

Como A es hermítica, es igual a A^H, por lo tanto

$$\begin{aligned} LDU &= U^H D L^H \\ D &= L^{-1}\big(U^H D L^H\big)U^{-1} \\ D &= \big(UL\big)^H D \big(UL\big)^H, \end{aligned}$$

luego $L = U^{-1} = U^H$, lo cual completa la prueba. ■

Ejemplo 4.8 *Determine la descomposición LDU para la matriz dada en el Ejemplo 4.6.*

Solución.

De manera análoga al caso real, se realizan las respectivas operaciones por filas y se llega a

$$\begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ i & 1 & 0 \\ -1 & -i & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -i & -1 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{bmatrix};$$

nótese que $L = U^H$.

Teorema 4.26 *Si N es una matriz normal, entonces la matriz $T = U^H N U$ (U unitaria) es también normal.*

Demostración.

Sea N una matriz normal y definamos $T = U^H N U$. Multiplicando por T^H, se obtiene que

$$\begin{aligned} TT^H &= \big(U^H N U\big)\big(U^H N U\big)^H \\ &= U^H N \big(UU^H\big) N^H U = U^H N N^H U \\ &= U^H N^H N U && \text{puesto que } N \text{ es normal} \\ &= \big(U^H N^H U\big)\big(U^H N U\big) = T^H T, \end{aligned}$$

y como $TT^H = T^H T$, se ha demostrado que T es normal. ■

Teorema 4.27 *Sea A una matriz de componentes complejas de tamaño $n \times n$. La matriz A es normal si y solo si es diagonalizable mediante una matriz unitaria.*

Demostración.

Supongamos que A es normal. Por el teorema de Schur, la matriz A es semejante a una matriz triangular superior T, por medio de una matriz unitaria U. Es decir $T = U^H AU$. Pero como

$$\begin{aligned} TT^H &= \left(U^H AU\right)\left(U^H AU\right)^H \\ &= U^H AA^H U = U^H A^H AU \\ &= \left(U^H A^H U\right)\left(U^H AU\right) = T^H T, \end{aligned}$$

la matriz T resulta ser normal. El lector puede verificar que la matriz T es diagonal.

Recíprocamente, supóngase que A es diagonalizable por una matriz unitaria U, es decir $U^H AU = D$, donde D es una matriz diagonal. Como D es normal (Ejemplo 4.3), se verifica que

$$\begin{aligned} AA^H &= \left(UDU^H\right)\left(UDU^H\right)^H \\ &= UDD^H U^H = UD^H DU^H \\ &= \left(UD^H U^H\right)\left(UDU^H\right) \\ &= A^H A. \end{aligned}$$

Por lo tanto, la matriz A es normal. ■

4.2.1 Forma canónica de Jordan

Si A es una matriz real de tamaño $n \times n$, en la Sección 3.4 se vio que se podía encontrar una matriz no singular P de tamaño $n \times n$ tal que

$$\mathcal{J} = P^{-1} AP.$$

En esta sección se explica, para matrices de componentes complejas, la forma canónica de Jordan.

Teorema 4.28 *Sea A una matriz compleja de tamaño $n \times n$. Entonces, existe una matriz P no singular tal que*

$$P^{-1}AP = \begin{bmatrix} \mathcal{J}_{n_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathcal{J}_{n_k}(\lambda_k) \end{bmatrix} = \mathcal{J}, \tag{4.15}$$

en donde cada $\mathcal{J}_{n_i}(\lambda_i)$ es un bloque de Jordan de tamaño $n_i \times n_i$ y $n_1 + n_2 + \ldots + n_k = n$. Los valores propios λ_i, $i = 1, 2, \ldots, k$ no son necesariamente distintos. El número total de bloques quedan determinados unívocamente por la matriz A.

Demostración.

Queda como ejercicio para el lector. ■

Ejemplo 4.9 *Encuentre una matriz no singular P tal que $P^{-1}AP$ sea una matriz de Jordan, para la siguiente matriz:*

$$\begin{bmatrix} 2+i & 1 & 0 \\ -2 & i & 1 \\ 1 & 1 & 1+i \end{bmatrix}.$$

Solución.

La ecuación característica de A es

$$-\lambda^3 + \big(3+3i\big)\lambda^2 - 6i\lambda - \big(2-2i\big) = 0.$$

Luego, $\lambda = 1 + i$ es el único valor propio (de multiplicidad algebraica tres). Entonces,

$$(A - \lambda I)\vec{v} = \big[A - (1+i)I\big]\vec{v} = \begin{bmatrix} 1 & 1 & 0 \\ -2 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Esto conduce a que $x + y = 0$ y $-x + z = 0$. Tomando $x \in \mathbb{R}$, se obtiene el vector propio: $\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. Para encontrar un vector propio generalizado, $\vec{v}_2$ se calcula

$$\begin{aligned} \big[A - (1+i)I\big]\vec{v}_2 &= \vec{v}_1 \\ \begin{bmatrix} 1 & 1 & 0 \\ -2 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Si se realizan operaciones por filas, se obtiene el vector propio generalizado: $\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, y de manera análoga, el vector propio generalizado: $\vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Por consiguiente,

$$P = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad \text{y} \qquad P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

Al efectuar el producto $P^{-1}AP$, se llega a

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2+i & 1 & 0 \\ -2 & i & 1 \\ 1 & 1 & 1+i \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+i & 1 & 0 \\ 0 & 1+i & 1 \\ 0 & 0 & 1+i \end{bmatrix}.$$

El lector puede notar que sobre la diagonal de la matriz de Jordan se encuentra el valor propio de la matriz A.

4.2.2 Descomposición en valores singulares

Si A es una matriz real de tamaño $m \times n$, hemos visto en la Sección 3.6 que se pueden encontrar dos matrices ortogonales U y V de tamaños $m \times m$ y $n \times n$, respectivamente, tales que

$$A = USV^t.$$

En esta sección se describe la descomposición en valores singulares y la descomposición polar para matrices de componentes complejas.

Teorema 4.29 Descomposición en valores singulares

Sea A una matriz compleja de tamaño $m \times n$ con rango r. Entonces, existen matrices unitarias U y V de tamaño $m \times m$ y $n \times n$, respectivamente, tales que

$$A = USV^H, \tag{4.16}$$

donde S es la matriz particionada de tamaño $m \times n$, dada por

$$S = \begin{bmatrix} D_{r\times r} & \vdots & \mathrm{O}_{r\times n_1} \\ \dots & \cdot & \dots \\ \mathrm{O}_{m_1\times r} & \vdots & \mathrm{O}_{m_1\times n_1} \end{bmatrix}, \quad \textit{donde} \quad D_{r\times r} = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix}, \tag{4.17}$$

siendo σ_i, para $i = 1, 2, \dots, r$, los valores singulares no nulos de A.

Demostración.

Queda como ejercicio para el lector. ■

Ejemplo 4.10 *Encuentre para la siguiente matriz de componentes complejas,*

$$A = \begin{bmatrix} 1+i & i \\ 2 & 1-i \end{bmatrix},$$

su descomposición en valores singulares.

Solución.

La matriz $A^H A$ es

$$\begin{bmatrix} 1-i & 2 \\ -i & 1+i \end{bmatrix} \begin{bmatrix} 1+i & i \\ 2 & 1-i \end{bmatrix} = \begin{bmatrix} 6 & 3-i \\ 3+i & 3 \end{bmatrix}.$$

En este caso, los valores propios de $A^H A$ son $\lambda_1 = 8$ y $\lambda_2 = 1$. Por lo tanto, los valores singulares asociados a la matriz A son $\sigma_1^2 = 8$ y $\sigma_2^2 = 1$. Al calcular los respectivos vectores propios normalizados, se obtiene

$$\vec{v}_1 = \frac{1}{\sqrt{35}} \begin{bmatrix} 5 \\ 3+i \end{bmatrix} \qquad \text{y} \qquad \vec{v}_2 = \frac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ -3-i \end{bmatrix}.$$

Por otra parte, la matriz AA^H es

$$\begin{bmatrix} 1+i & i \\ 2 & 1-i \end{bmatrix} \begin{bmatrix} 1-i & 2 \\ -i & 1+i \end{bmatrix} = \begin{bmatrix} 3 & 1+3i \\ 1-3i & 6 \end{bmatrix}$$

y sus respectivos vectores propios normalizados son

$$\vec{u}_1 = \frac{1}{\sqrt{70}} \begin{bmatrix} 2+4i \\ 7-i \end{bmatrix} \qquad \text{y} \qquad \vec{u}_2 = \frac{1}{\sqrt{14}} \begin{bmatrix} 3-i \\ 2i \end{bmatrix}.$$

Finalmente, si $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix}$, $V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$ y $S = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & 1 \end{bmatrix}$, entonces la SVD de A es

$$A = \left(\frac{1}{\sqrt{7}}\right)^2 \begin{bmatrix} \frac{2+4i}{\sqrt{10}} & \frac{3-i}{\sqrt{2}} \\ \frac{7-i}{\sqrt{10}} & \frac{2i}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{5}} & \frac{3-i}{\sqrt{5}} \\ \frac{2}{\sqrt{2}} & \frac{-3+i}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1+i & i \\ 2 & 1-i \end{bmatrix}.$$

4.2.3 Descomposición polar

Ahora se estudiará la *descomposición polar* para matrices de componentes complejas. El nombre de descomposición polar se debe a la representación polar de un número complejo $z = \rho e^{i\theta}$. La analogía entre esta representación de los números complejos y la descomposición (3.42) de una matriz es debida a que los valores propios de la matriz P son números reales no negativos y los de la matriz Q son números complejos unitarios.

Teorema 4.30 Descomposición polar a izquierda

Sea A una matriz compleja de tamaño $n \times n$ con rango r. Entonces, existe una matriz hermitiana P de tamaño $n \times n$ con valores propios no negativos y una matriz unitaria Q de tamaño $n \times n$ tales que

$$A = PQ. \tag{4.18}$$

Demostración.

Si A es una matriz de tamaño $n \times n$, las matrices U, S y V de la descomposición en valores singulares son también de tamaño $n \times n$. Luego, la ecuación (4.16) se puede escribir como

$$A = USV^H = US(U^HU)V^H = (USU^H)UV^H = PQ.$$

La matriz P es siempre única, aun cuando A sea singular, y está dada por

$$P = \left[AA^H\right]^{\frac{1}{2}} = USU^H.$$

Nótese que esta matriz es hermítica, ya que

$$\begin{aligned} P^H &= \left(USU^H\right)^H = US^HU^H \\ &= USU^H = P \qquad\qquad \text{(por ser } S \text{ simétrica)}, \end{aligned}$$

y la matriz $Q = UV^H$ es unitaria, puesto que

$$Q^{-1} = \left(UV^H\right)^{-1} = \left(V^H\right)^{-1}U^{-1} = VU^H = Q^H.$$

En la última ecuación se usaron los hechos de que U y V eran matrices unitarias.

Se deja como ejercicio la comprobación de que P tiene valores propios no negativos. ■

Corolario 4.30.1 Descomposición polar a derecha

Sea A una matriz compleja de tamaño $n \times n$ con rango r. Entonces, existe una matriz hermitiana P' de tamaño $n \times n$ con valores propios no negativos y una matriz unitaria Q de tamaño $n \times n$ tales que

$$A = QP'. \tag{4.19}$$

Demostración.

En términos de la descomposición en valores singulares de A, es decir $A = USV^H$, se tiene que

$$P' = \left[A^HA\right]^{\frac{1}{2}} = VSV^H \qquad \text{y} \qquad Q = UV^H.$$

Si A es no singular, entonces la matriz Q está dada por

$$Q = A\left[P'\right]^{-1}.$$

El lector puede verificar que P' tiene valores propios no negativos y Q es unitaria. ■

La descomposición de la conjugada compleja de A está dada por

$$\overline{A} = \overline{Q}\,\overline{P'}.$$

Esta descomposición siempre existe, y cuando la matriz A es no singular, entonces es única y la matriz P' tiene valores propios positivos. Por otra parte,

$$\det(A) = \det\big(P'\big)\ \det(Q) = \rho e^{i\theta}$$

es la respectiva descomposición polar del determinante de A, ya que

$$\det\big(P'\big)\ = \rho = \big|\det(A)\big| \qquad \text{y} \qquad \det(Q)\ = e^{i\theta}.$$

La descomposición polar a izquierda es también conocida como la descomposición polar inversa. Las descomposiciones polares a izquierda y a derecha están relacionadas por

$$P\ = QP'Q^H \qquad \text{con} \qquad Q\ = UV^H.$$

La matriz A es una matriz normal si y solo si $P' = P$. En este caso, $\widetilde{Q}S = S\widetilde{Q}$, donde $\widetilde{Q} = V^H U$ y por el Teorema 4.23, es posible diagonalizarlas mediante una matriz unitaria R, luego $\widetilde{Q} = R^H \Delta R$, donde Δ es una matriz diagonal unitaria de fase $e^{i\varphi}$. Al tomar $W = VR^H$, se puede volver a escribir la descomposición polar como

$$A = \big(W\Delta W^H\big)\big(WSW^H\big),$$

por lo tanto, A también tiene una descomposición espectral dada por

$$A = W\Lambda W^H,$$

con valores propios complejos tales que $\Lambda\Lambda^H = S^2$ y una matriz unitaria W de vectores propios complejos.

Ejemplo 4.11 *Encuentre para la matriz dada en el Ejemplo 4.10 la descomposición polar tanto a izquierda como a derecha.*

Solución.

La matriz P es

$$USU^H = \frac{1}{7}\begin{bmatrix} 4\sqrt{2}+5 & \big(2+6i\big)\sqrt{2}-1-3i \\ \big(2-6i\big)\sqrt{2}-1+3i & 10\sqrt{2}+2 \end{bmatrix}$$

y la matriz Q unitaria es

$$UV^H = \frac{1}{7}\begin{bmatrix} (1+2i)\sqrt{2}+3-i & (1+i)\sqrt{2}-4+3i \\ \left(\frac{7}{2}-\frac{1}{2}i\right)\sqrt{2}+2i & (2-i)\sqrt{2}-1-3i \end{bmatrix}.$$

El lector puede verificar que $A = PQ$. Para determinar la descomposición polar a derecha, se establece

$$P' = VSV^H = \left(\frac{1}{\sqrt{7}}\right)^2 \begin{bmatrix} 10\sqrt{2}+2 & (6-2i)\sqrt{2}-3+i \\ (6+2i)\sqrt{2}-3-i & 4\sqrt{2}+5 \end{bmatrix}.$$

Luego, la descomposición polar a derecha queda $A = QP'$.

Las propiedades de las matrices complejas descritas en este capítulo son comparables a las propiedades de las matrices reales analizadas anteriormente. En el siguiente resumen se indica la correspondencia entre las matrices complejas unitarias y hermitianas con las matrices reales ortogonales y simétricas.

Comparación entre matrices reales y matrices complejas

Sea $A = [a_{ij}]$, con $a_{ij} \in \mathbb{R}$.	Sea $A = [a_{ij}]$, con $a_{ij} \in \mathbb{C}$.
1. Toda matriz simétrica $(A = A^t)$ tiene valores propios reales.	1. Toda matriz hermitiana $(A = A^H)$ tiene valores propios reales.
2. Si A es una matriz simé-trica, los vectores propios correspondientes a valores propios distintos son ortogonales.	2. Si A es una matriz hermitiana, los vectores propios correspondientes a valores propios distintos son ortogonales.
3. *Descomposición de Schur.* Si A es una matriz de tamaño $n \times n$ con valores propios reales, existe una matriz ortogonal Q tal que $$Q^t AQ = T$$ es una matriz triangular superior.	3. *Descomposición de Schur.* Si A es una matriz de tamaño $n \times n$, existe una matriz unitaria U tal que $$U^H AU = T$$ es una matriz triangular superior.
4. *Teorema espectral.* Si $A = A^t$, existe una matriz ortogonal Q tal que $$Q^t AQ = D$$ es una matriz diagonal.	4. *Teorema espectral.* Si $A = A^H$, existe una matriz unitaria U tal que $$U^H AU = D.$$ es una matriz diagonal.

5. *Descomposición en valores singulares.* Existen matrices ortogonales U y V de tamaños $m \times m$ y $n \times n$, respectivamente, tales que $$A = USV^t,$$ donde S está dada por (3.39). 6. *Descomposición polar.* Existe una matriz simétrica P de tamaño $n \times n$ con valores propios no negativos y una matriz ortogonal Q de tamaño $n \times n$ tal que $$A = PQ,$$ donde $P = USU^t$ y $Q = UV^t$.	5. *Descomposición en valores singulares.* Existen matrices unitarias U y V de tamaños $m \times m$ y $n \times n$, respectivamente, tales que $$A = USV^H,$$ donde S está dada por (4.17). 6. *Descomposición polar.* Existe una matriz hermitiana P de tamaño $n \times n$ con valores propios no negativos y una matriz unitaria Q de tamaño $n \times n$ tal que $$A = PQ,$$ donde $P = USU^H$ y $Q = UV^H$.

Ejercicios 4.1

1. Determine para cada una de las siguientes matrices una matriz unitaria U tal que $U^H AU$ sea diagonal:

$a.$ $\begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}$. $b.$ $\begin{bmatrix} 1 & 2+3i \\ 2-3i & -1 \end{bmatrix}$. $c.$ $\begin{bmatrix} 2 & 1+2i \\ 1-2i & -2 \end{bmatrix}$.

$d.$ $\begin{bmatrix} 1 & i & 2+i \\ -i & 2 & 1-i \\ 2-i & 1+i & 2 \end{bmatrix}$. $e.$ $\begin{bmatrix} 1 & 1+i & 2i \\ 1-i & 4 & 2-3i \\ -2i & 2+3i & 7 \end{bmatrix}$.

2. Sea $\mathbb{V} = \{A \in M_{22} : a_{ij} \in \mathbb{C}\}$. Determine si cada uno de los siguientes subconjuntos $\mathbb{H}$ son o no subespacios de $\mathbb{V}$:

   a) $\mathbb{H} = \{A \in \mathbb{V} : a_{ii} = 0\}$.

b) $\mathbb{H} = \{A \in \mathbb{V} : a_{ii} \text{ son imaginarios puros}\}$.

c) $\mathbb{H} = \left\{A \in \mathbb{V} : A = A^H\right\}$.

3. Sea $\mathbb{H} = \left\{ A \in M_{22} : A = \begin{bmatrix} w & -\overline{z} \\ \overline{z} & w \end{bmatrix}; w, z \in \mathbb{C} \right\}$.

a) Demuestre que $\mathbb{H}$ es cerrado para la suma y la multiplicación.

b) ¿Cuáles matrices en $\mathbb{H}$ son no singulares?

c) Compruebe que si una matriz en $\mathbb{H}$ es no singular, entonces la inversa está en $\mathbb{H}$.

d) Encuentre dos matrices A y B en $\mathbb{H}$ tal que $AB \neq BA$.

4. Sea A una matriz de tamaño $n \times n$ con componentes complejas y sea $\vec{x} \in \mathbb{C}^n$ un vector propio correspondiente al valor propio $\lambda \in \mathbb{C}$. Muestre que para cada escalar complejo no nulo α el vector $\alpha\vec{x}$ es un vector propio de A.

5. Si A es una matriz normal, pruebe que A y A^H son diagonalizables por la misma matriz unitaria.

6. Si A es una matriz normal, pruebe que λ es un valor propio de A si y solo si $\overline{\lambda}$ es un valor propio de A^H.

Capítulo 5

Formas bilineales

En este capítulo estudiaremos las formas bilineales sobre espacios de dimensión finita. Se introduce la representación matricial de una forma bilineal y se establece el isomorfismo entre el espacio de las formas y el espacio de las matrices de tamaño $n \times n$.

5.1 Formas bilineales

Definición 5.1 *Sean $\mathbb{U}, \mathbb{V}$ y $\mathbb{W}$ espacios vectoriales reales. Una aplicación $g : \mathbb{U} \times \mathbb{V} \to \mathbb{W}$ se llama bilineal si satisface las siguientes propiedades:*

BI 1. *Para todo $\vec{u}_1, \vec{u}_2 \in \mathbb{U}$ y $\vec{v} \in \mathbb{V}$, se tiene que*

$$g\big(\vec{u}_1 + \vec{u}_2, \vec{v}\big) = g\big(\vec{u}_1, \vec{v}\big) + g\big(\vec{u}_2, \vec{v}\big),$$

y para todo $\vec{u} \in \mathbb{U}$ y $\vec{v}_1, \vec{v}_2 \in \mathbb{V}$, se tiene que

$$g\big(\vec{u}, \vec{v}_1 + \vec{v}_2\big) = g\big(\vec{u}, \vec{v}_1\big) + g\big(\vec{u}, \vec{v}_2\big).$$

BI 2. *Para todo $\alpha \in \mathbb{R}, \vec{u} \in \mathbb{U}$ y $\vec{v} \in \mathbb{V}$, se tiene que*

$$g\big(\alpha\vec{u}, \vec{v}\big) = \alpha g\big(\vec{u}, \vec{v}\big) \ = g\big(\vec{u}, \alpha\vec{v}\big).$$

Ejemplo 5.1 *Sea* $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ *definida por*

$$g(\vec{X}, \vec{Y}) = \vec{X}^t A \vec{Y},$$

donde $\vec{X} \in \mathbb{R}^m$, $\vec{Y} \in \mathbb{R}^n$ *y* A *es una matriz real de tamaño* $m \times n$. *Verifique si la aplicación* g *es bilineal.*

Solución.

Para todo $\vec{X}_1,\ \vec{X}_2 \in \mathbb{R}^m$ y $\vec{Y} \in \mathbb{R}^n$ se tiene que

$$\begin{aligned} g(\vec{X}_1 + \vec{X}_2, \vec{Y}) &= (\vec{X}_1 + \vec{X}_2)^t A \vec{Y} = (\vec{X}_1^t + \vec{X}_2^t) A \vec{Y} \\ &= \vec{X}_1^t A \vec{Y} + \vec{X}_2^t A \vec{Y} = g(\vec{X}_1, \vec{Y}) + g(\vec{X}_2, \vec{Y}). \end{aligned}$$

Para todo $\alpha \in \mathbb{R}$, $\vec{X} \in \mathbb{R}^m$ y $\vec{Y} \in \mathbb{R}^n$ se tiene que

$$\begin{aligned} g(\alpha \vec{X}, \vec{Y}) &= (\alpha \vec{X})^t A \vec{Y} = (\alpha^t \vec{X}^t) A \vec{Y} \\ &= \alpha \vec{X}^t A \vec{Y} = \alpha g(\vec{X}, \vec{Y}). \end{aligned}$$

Así, la aplicación g es lineal cuando $\vec{Y} \in \mathbb{R}^n$ permanece fijo. De manera análoga, se puede probar que g es una transformación lineal cuando la componente $\vec{X} \in \mathbb{R}^m$ se mantiene fija. Por lo tanto, g es una aplicación bilineal.

Teorema 5.1 *Sea* $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ *una aplicación bilineal. Entonces existe una matriz única* A *de tamaño* $m \times n$, *tal que*

$$g(\vec{X}, \vec{Y}) = g_A(\vec{X}, \vec{Y}) = \vec{X}^t A \vec{Y}. \tag{5.1}$$

El conjunto de aplicaciones bilineales de $\mathbb{R}^m \times \mathbb{R}^n$ *en* $\mathbb{R}$ *es un espacio vectorial denotado por* $Bil(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$, *y la asociación*

$$A \mapsto g_A$$

es un isomorfismo entre $Bil(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ *y el espacio de las matrices reales de tamaño* $m \times n$.

Demostración.

Sean $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_m$ los vectores unitarios estándar para $\mathbb{R}^m$ y sean $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ los vectores unitarios estándar para $\mathbb{R}^n$. Luego, se puede expresar cualquier $\vec{X} \in \mathbb{R}^m$ y cualquier $\vec{Y} \in \mathbb{R}^n$ de la siguiente manera

$$\vec{X} = \sum_{i=1}^{m} x_i \vec{e}_i \qquad \text{y} \qquad \vec{Y} = \sum_{j=1}^{n} y_j \vec{u}_j.$$

Entonces, se tiene que

$$g\big(\vec{X}, \vec{Y}\big) = g\left(\sum_{i=1}^{m} x_i \vec{e}_i, \sum_{j=1}^{n} y_j \vec{u}_j \right).$$

Como g es una transformación lineal, en la primera componente se llega a

$$g\big(\vec{X}, \vec{Y}\big) = \sum_{i=1}^{m} x_i \left[g\left(\vec{e}_i, \sum_{j=1}^{n} y_j \vec{u}_j \right) \right],$$

y usando el hecho de que g es lineal, en la segunda componente se halla que

$$g\big(\vec{X}, \vec{Y}\big) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j \; g\big(\vec{e}_i, \vec{u}_j\big).$$

Sea

$$a_{ij} = g\big(\vec{e}_i, \vec{u}_j\big).$$

Luego, (5.1) se puede expresar como sigue:

$$g\big(\vec{X}, \vec{Y}\big) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j, \tag{5.2}$$

que es precisamente la expresión del producto

$$\vec{X}^t A \vec{Y} \qquad \text{con} \qquad A = [a_{ij}]\,;$$

esto prueba que $g = g_A$ para las a_{ij} escogidas anteriormente.

Ahora se demuestra que A es única, para ello, suponga que $g\big(\vec{X}, \vec{Y}\big) = \vec{X}^t A \vec{Y}$ y que $g\big(\vec{X}, \vec{Y}\big) = \vec{X}^t B \vec{Y}$ para todo $\vec{X} \in \mathbb{R}^m$ y $\vec{Y} \in \mathbb{R}^n$. Entonces,

$$\vec{X}^t A \vec{Y} = \vec{X}^t B \vec{Y},$$

y al establecer $C = A - B$, se tiene que

$$\vec{X}^t C\vec{Y} = 0, \qquad \text{para todo} \qquad \vec{X} \in \mathbb{R}^m \text{ y } \vec{Y} \in \mathbb{R}^n.$$

En particular, si $\vec{X} = \vec{e}_i$ e $\vec{Y} = \vec{u}_j$, se tiene que

$$0 = \vec{e}_i{}^t C\vec{u}_j = c_{ij}.$$

Así que $c_{ij} = 0$ para todo $i,\ j$, y por lo tanto $C = \mathrm{O}$ es la matriz cero de tamaño $m \times n$. Esto muestra que $A = B$.

El lector puede probar como ejercicio la parte referente al isomorfismo entre el espacio de las matrices y las aplicaciones bilineales. ■

Definición 5.2 *La matriz A en el Teorema 5.1 se llama representación matricial de la aplicación bilineal g_A.*

Definición 5.3 Forma bilineal

Si en la definición 5.1 se tiene que los espacios $\mathbb{U} = \mathbb{V}$ y el espacio $\mathbb{W} = \mathbb{R}$ de tal manera que g aplica a $\mathbb{V} \times \mathbb{V}$ en $\mathbb{R}$, entonces se dice que g es una forma bilineal sobre $\mathbb{V}$.

Teorema 5.2 *Sean $g_1,\ g_2 : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ dos formas bilineales distintas sobre $\mathbb{V}$. Entonces:*

a) $g_1 + g_2$ es una forma bilineal.

b) αg_1 es también una forma bilineal, donde $\alpha \in \mathbb{R}$.

Demostración.
Queda como ejercicio para el lector. ■

Definición 5.4 Rango de una forma bilineal

El rango de una forma bilineal g sobre $\mathbb{V}$, escrito $\rho(g)$, se define como el rango de la matriz que representa a g. Se dice que g es degenerada o no degenerada según si $\rho(g) < \dim(\mathbb{V})$ o $\rho(g) = \dim(\mathbb{V})$.

Definición 5.5 *Sea $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ una forma bilineal sobre $\mathbb{V}$. Entonces g es simétrica si para todo $\vec{v},\ \vec{w} \in \mathbb{V}$, se cumple que*

$$g\big(\vec{v}, \vec{w}\big) = g\big(\vec{w}, \vec{v}\big). \tag{5.3}$$

Teorema 5.3 *Una matriz real A de tamaño $n \times n$ representa una forma bilineal simétrica si y solo si es una matriz simétrica.*

Demostración.

Supóngase que A es simétrica. Como para todo $\vec{X},\ \vec{Y} \in \mathbb{R}^n$, la matriz $\vec{X}^t A\vec{Y}$ es una matriz de 1×1, es decir, un elemento de $\mathbb{R}$, entonces es igual a su propia transpuesta. Por lo tanto,

$$\vec{X}^t A\vec{Y} = (\vec{X}^t A\vec{Y})^t = \vec{Y}^t A^t (\vec{X}^t)^t = \vec{Y}^t A\vec{X}.$$

Así que A representa una forma bilineal simétrica.

Recíprocamente, supóngase que A representa una forma bilineal simétrica, es decir,

$$g_A\big(\vec{X}, \vec{Y}\big) = g_A\big(\vec{Y}, \vec{X}\big) \tag{5.4}$$

para todo $\vec{X},\ \vec{Y} \in \mathbb{R}^n$. Como

$$g_A\big(\vec{Y}, \vec{X}\big) = \vec{Y}^t A\vec{X} = \big(\vec{Y}^t A\vec{X}\big)^t = \vec{X}^t A^t \big(\vec{Y}^t\big)^t = \vec{X}^t A^t \vec{Y}, \tag{5.5}$$

si se comparan las expresiones (5.4) y (5.5), se tiene que

$$g_A\big(\vec{X}, \vec{Y}\big) = \vec{X}^t A\vec{Y} = \vec{X}^t A^t \vec{Y}. \tag{5.6}$$

Como (5.6) se cumple para todo $\vec{X},\ \vec{Y} \in \mathbb{R}^n$, se concluye que $A = A^t$. Es decir, A es simétrica. ■

Definición 5.6 *Sea $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ una forma bilineal sobre $\mathbb{V}$. Entonces g es alternada si para todo $\vec{v},\ \vec{w} \in \mathbb{V}$, se cumple que*

1. $g\big(\vec{v}, \vec{v}\big) = 0,$

2. $g\big(\vec{v}, \vec{w}\big) = -\,g\big(\vec{w}, \vec{v}\big).$

Definición 5.7 Inercia

Sea A una matriz simétrica de tamaño $n \times n$. La inercia de A es la terna ordenada de números:

$$In(A) = \big(pos, neg, nul\big), \tag{5.7}$$

donde pos, neg y nul son los números de valores propios de A positivos, negativos y nulos, respectivamente (contando todas las multiplicidades algebraicas).

Nótese que $\rho(A) = pos + neg$.

Definición 5.8 Signatura

A la diferencia entre el número de valores propios positivos y el número de valores propios negativos se le denomina signatura de la matriz A. En otras palabras, si $In(A) = (i, j, k)$, se llama signatura de la matriz A a la cantidad:

$$Sig(A) = i - j.$$

Ejemplo 5.2 *Determinar la inercia y signatura de la matriz dada en el Ejemplo 2.18.*

Solución.
De los resultados del Ejemplo 2.18 se tiene que los valores propios de la matriz A eran $\lambda_1 = 1$ de multiplicidad algebraica 2 y $\lambda_2 = 7$. Luego,

$$In(A) = (3, 0, 0) \qquad \text{y} \qquad Sig(A) = 3.$$

Ejercicios 5.1

1. Asumiendo $\vec{X}^t = [x_1 \;\; x_2]$ y $\vec{Y}^t = [y_1 \;\; y_2]$, determine cuáles de las siguientes funciones $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ son aplicaciones bilineales:

   a. $g(\vec{X}, \vec{Y}) = 4x_1y_1 + 2x_2y_1 + 2x_1y_2 + x_2y_2$.

   b. $g(\vec{X}, \vec{Y}) = 9x_1y_1 - 12x_1y_2 - 12x_2y_1 + 16x_2y_2$.

   c. $g(\vec{X}, \vec{Y}) = x_1y_1 + 4x_1y_2 + 4x_2y_1 + 16x_2y_2$.

   d. $g(\vec{X}, \vec{Y}) = x_1y_1 + 2x_1y_2 - 2x_2y_1 + x_2y_2$.

   e. $g(\vec{X}, \vec{Y}) = x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2$.

   f. $g(\vec{X}, \vec{Y}) = x_1y_2 - x_2y_1$.

2. Escriba cada una de las siguientes formas bilineales $F : \mathbb{R}^3 \to \mathbb{R}$ en forma matricial:

   a. $F(x, y, z) = 6x^2 + 3y^2 - 2z^2 + 12xy - 18yz$.

   b. $F(x, y, z) = 7x^2 + 7y^2 + 10z^2 - 2xy - 4xz + 4yz$.

   c. $F(x, y, z) = 2x^2 + 2y^2 + 5z^2 - 4xy + 2xz - 2yz$.

3. Sea $\mathbb{V} = M_{23}$. Para $A,\ B \in \mathbb{V}$, defina $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ como

$$g(A, B) = \operatorname{tr}(A^tB).$$

   Demuestre que es una aplicación bilineal en $\mathbb{V} \times \mathbb{V}$.

4. Sea $\mathbb{V} = M_{nn}$, demuestre que la función $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ dada por

$$g(A, B) = n \operatorname{tr}(AB) - \operatorname{tr}(A) \operatorname{tr}(B)$$

   es una aplicación bilineal en $\mathbb{V} \times \mathbb{V}$.

5. Si $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ y $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, represéntese la forma bilineal

$$Cov\big(\vec{X},\vec{Y}\big) = \frac{1}{n-1}\sum_{i=1}^{n}\big(x_i-\overline{x}\big)\big(y_i-\overline{y}\big)$$

en la forma $\vec{X}^t A\vec{Y}$, con A simétrica. ¿Cuál es el rango de A?

5.2 Formas cuadráticas

Cuando se considera el cuadrado de la norma de un vector $\vec{X} \in \mathbb{R}^n$, se obtiene la expresión $\|\vec{X}\|^2 = \vec{X}^t\vec{X}$; tales sumas y expresiones en forma general se denominan *formas cuadráticas*. Ellas surgen frecuentemente en una gran variedad de aplicaciones. Por ejemplo, se pueden usar formas cuadráticas en ingeniería (para optimización), en economía (en el análisis de funciones de costo y utilidad), en física (para el estudio de energías cinéticas y potenciales) y en estadística (en el análisis de varianza). En esta sección se estudiarán algunos temas relacionados con estas formas, utilizando la teoría de las matrices simétricas analizada anteriormente.

Definición 5.9 *Sea $\mathbb{V}$ un espacio vectorial real de dimensión finita. Sea $g : \mathbb{V}\times\mathbb{V} \to \mathbb{R}$ una forma bilineal simétrica sobre $\mathbb{V}$. Entonces, una forma cuadrática determinada por g es una función $F : \mathbb{V} \to \mathbb{R}$ tal que*

$$F\big(\vec{v}\big) = g_A\big(\vec{v},\vec{v}\big) = \vec{v}^t A\vec{v}. \tag{5.8}$$

La matriz A es llamada matriz de la forma cuadrática.

Ejemplo 5.3 *Sean $\mathbb{V} = \mathbb{R}^n$ y $F\big(\vec{v}\big) = v_1^2 + v_2^2 + \ldots + v_n^2$. Escriba esta forma cuadrática como $\vec{v}^t A\vec{v}$.*

Solución.

Usando la expresión (5.2) para $m = n$, se puede determinar la matriz $A = [a_{ij}]$ asociada a la forma bilineal simétrica g, es decir,

$$F(\vec{v}) = \vec{v}^t A\vec{v} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}v_iv_j. \tag{5.9}$$

En otras palabras, se quiere encontrar los valores de a_{ij}, de manera que

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}v_iv_j = \sum_{i=1}^{n} v_i^2.$$

Como la matriz A es simétrica $a_{ij} = a_{ji}$, la forma cuadrática dada en (5.9) se puede expresar como sigue:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}v_iv_j = \sum_{i=1}^{n} a_{ii}v_i^2 + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} a_{ij}v_iv_j. \tag{5.10}$$

Si se comparan los términos, se establecen las siguientes relaciones:

$$\sum_{i=1}^{n} a_{ii}v_i^2 = \sum_{i=1}^{n} v_i^2 \qquad \text{y} \qquad 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} a_{ij}v_iv_j = 0.$$

Pero como en la función $F(\vec{v})$ no aparecen términos de la forma v_iv_j, entonces

$$a_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Luego, $A = I_n$ y, por lo tanto, $F(\vec{v})$ se puede expresar como $\vec{v}\,{}^t I_n\vec{v}$.

Ejemplo 5.4 *Sean* $\mathbb{V} = \mathbb{R}^3$ *y* $F(\vec{X}) = 2x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 + 4x_2x_3$. *Exprese esta forma cuadrática como* $\vec{X}^t A\vec{X}$.

Solución.

Utilizando el resultado obtenido en (5.10) para $n = 3$, se tiene que

$$\vec{X}^t A\vec{X} = \sum_{i=1}^{3} a_{ii}x_i^2 + 2\sum_{i=1}^{2}\sum_{j>i}^{3} a_{ij}x_ix_j.$$

Si se desarrolla esta suma y se comparan los a_{ij} con los coeficientes de la función $F(\vec{X})$, se obtiene la matriz

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix},$$

la cual permite expresar a $F(\vec{X})$ de la forma $\vec{X}^t A \vec{X}$.

Ejemplo 5.5 *Si* $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, *represéntese la forma cuadrática*

$$(n-1)s_x^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

en la forma $\vec{X}^t A \vec{X}$, *con* A *simétrica. ¿Cuál es el rango de* A*?*

Solución.

Sea $\vec{X}^t = (x_1, x_2, \ldots, x_n)$ el vector que representa las n-observaciones. Si en el Ejemplo 5.3 se reemplaza cada v_i por $x_i - \overline{x}$, se tiene que

$$F(\vec{v}) = \sum_{i=1}^{n} (x_i - \overline{x})^2 = \vec{v}^t I_n \vec{v}; \tag{5.11}$$

aquí $\vec{v}$ se puede reescribir como

$$\vec{v} = \begin{bmatrix} x_1 - \overline{x} \\ x_2 - \overline{x} \\ \vdots \\ x_j - \overline{x} \\ \vdots \\ x_n - \overline{x} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix} - \overline{x} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \vec{X} - \mathbf{1}\overline{x}, \tag{5.12}$$

donde $\mathbf{1}$ es un vector columna de unos de tamaño $n \times 1$. Luego, $\overline{x}$ se puede expresar como sigue:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{1}^t \vec{X}. \tag{5.13}$$

Si se reemplaza (5.13) en (5.12), se obtiene

$$\vec{v} = \vec{X} - \frac{1}{n}\mathbf{1}\mathbf{1}^t \vec{X} = \left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\vec{X}. \tag{5.14}$$

Al sustituir en (5.11), se tiene que

$$\begin{aligned} F\Big[\Big(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\Big)\vec{X}\Big] &= \Big[\Big(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\Big)\vec{X}\Big]^t I_n\Big[\Big(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\Big)\vec{X}\Big] \\ &= \vec{X}^t\Big(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\Big)\Big(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\Big)\vec{X} \\ &= \vec{X}^t\Big(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^t + \frac{1}{n^2}\mathbf{1}\underbrace{\mathbf{1}^t\mathbf{1}}\mathbf{1}^t\Big)\vec{X} \\ &= \vec{X}^t\Big(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\Big)\vec{X}. \end{aligned}$$

Aquí se usó que $\mathbf{1}^t\mathbf{1} = n$, denotando $J_n = \mathbf{1}\mathbf{1}^t$, se llega finalmente a que

$$(n-1)s_x^2 = \vec{X}^t\Big(I_n - \frac{1}{n}J_n\Big)\vec{X} = \vec{X}^t(I_n - \overline{J}_n)\vec{X}. \tag{5.15}$$

Luego, la matriz asociada a la forma es

$$A = I_n - \overline{J}_n \qquad \text{y} \qquad \rho(A) = n - 1.$$

En este material, la matriz J_n siempre denotará la matriz con todos sus elementos iguales a uno de tamaño $n \times n$, definida anteriormente.

Definición 5.10 Forma polar de una forma cuadrática

Dada F una forma cuadrática real, se puede obtener una forma bilineal simétrica g de F de acuerdo con la siguiente identidad llamada la forma polar de g:

$$g(\vec{u}, \vec{v}) = \frac{1}{2}\big[F(\vec{u} + \vec{v}) - F(\vec{u}) - F(\vec{v})\big]. \tag{5.16}$$

Definición 5.11 *Dos formas cuadráticas $\vec{X}^t A\vec{X}$ y $\vec{Y}^t B\vec{Y}$ se dice que son equivalentes si existe una matriz no singular P tal que $B = P^t AP$. Aún más, las formas son equivalentes ortogonalmente si P se puede escoger ortogonal, equivalente-real si P se puede escoger con elementos reales y equivalente-compleja, o simplemente equivalente, si P tiene elementos complejos.*

Teorema 5.4 *Dos formas cuadráticas $\vec{X}^t A\vec{X}$ y $\vec{Y}^t B\vec{Y}$ son equivalentes si y solo si las matrices simétricas A y B son congruentes.*

Demostración.

Si en la forma cuadrática $\vec{X}^t A\vec{X}$ se hace el cambio de variable $\vec{X} = P\vec{Y}$, donde P es una matriz no singular, se obtiene la forma

$$\vec{X}^t A\vec{X} = \vec{Y}^t P^t AP\vec{Y} = \vec{Y}^t B\vec{Y}.$$

Recíprocamente, sean $A = [a_{ij}]$ y $P^t AP = [b_{ij}]$, las dos matrices simétricas asociadas con las formas cuadráticas, donde P es una matriz real no singular, entonces

$$\sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} a_{ij}x_i x_j \qquad \text{y} \qquad \sum_{i=1}^{n} b_{ii}y_i^2 + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} b_{ij}y_i y_j.$$

El cambio de variable $\vec{X} = P\vec{Y}$ cambia la primera forma cuadrática a la segunda. ■

Ejemplo 5.6 *Muestre que las formas cuadráticas:*

$$F_1(\vec{X}) = x_1^2 + x_2^2 + 8x_1x_2 \qquad y \qquad F_2(\vec{Y}) = y_1^2 - 14y_2^2 + 2y_1y_2$$

son equivalentes.

Solución.

Utilizando el resultado obtenido en (5.10), para $n = 2$, se tiene que

$$\vec{X}^t A\vec{X} = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2.$$

Si se comparan los a_{ij} con los coeficientes de la función $F_1(\vec{X})$, se obtiene la matriz

$$A = \begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix},$$

la cual permite expresar a $F_1(\vec{X})$ de la forma $\vec{X}^t A\vec{X}$. Para la forma cuadrática $F_2(\vec{Y})$, se tiene la matriz

$$B = \begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix}.$$

En el Ejemplo 2.8 se mostró que A y B eran congruentes. Por lo tanto, $F_1(\vec{X})$ es equivalente a la forma cuadrática $F_2(\vec{Y})$.

Teorema 5.5 *Dos formas cuadráticas $\vec{X}^t A\vec{X}$ y $\vec{Y}^t B\vec{Y}$ son equivalentes ortogonalmente si y solo si A y B tienen los mismos valores propios y estos ocurren con la misma multiplicidad.*

Demostración.

Si A y B tienen valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$ y D es una matriz diagonal con $\lambda_1, \lambda_2, \ldots, \lambda_n$ como elementos de su diagonal, entonces existen matrices ortogonales P y Q tal que

$$Q^t AQ = P^t BP = D.$$

Por consiguiente, $B = (P^t)^{-1}[Q^t AQ]P^{-1} = (QP^{-1})^t A(QP^{-1})$, y como QP^{-1} es ortogonal, $\vec{Y}^t B\vec{Y}$ es ortogonalmente equivalente a $\vec{X}^t A\vec{X}$.

Recíprocamente, si las dos formas son ortogonalmente equivalentes, B es similar a A (porque $P^{-1} = P^t$), y A y B tienen los mismos valores propios con las mismas multiplicidades. ■

Ejercicios 5.2

1. Exprese las siguientes formas cuadráticas de la forma $\vec{X}^t A\vec{X}$:

   a. $F(x, y, z) = 6x^2 + 3y^2 - 2z^2 + 12xy - 18yz.$

   b. $F(x, y, z) = 7x^2 + 7y^2 + 10z^2 - 2xy - 4xz + 4yz.$

   c. $F(x, y, z) = 2x^2 + 2y^2 + 5z^2 - 4xy + 2xz - 2yz.$

2. Determine si las siguientes formas cuadráticas son equivalentes.

   a. $F_1(\vec{X}) = 4x^2 + 4xy + y^2$ y $F_2(\vec{Y}) = 16u^2 - 24uv + 9v^2.$

   b. $F_1(\vec{X}) = x^2 + 8xy + 16y^2$ y $F_2(\vec{Y}) = u^2 + 2uv + v^2.$

3. Demuestre que la forma cuadrática $F : \mathbb{R}^2 \to \mathbb{R}$ dada por

   $$F(x, y) = ax^2 + 2bxy + cy^2$$

   tiene rango 2 si y solo si $ac - b^2 \neq 0$.

4. Determine los valores de α para los cuales la matriz asociada a la forma $F : \mathbb{R}^3 \to \mathbb{R}$ dada por

$$F(x, y, z) = 5x^2 + y^2 + \alpha z^2 + 4xy - 2xz - 2yz$$

tiene valores propios positivos.

5. Sea A una matriz simétrica de tamaño $n \times n$ asociada a la forma cuadrática $F(\vec{X}) = \vec{X}^H A\vec{X}$ con $\vec{X} \in \mathbb{C}^n$. Demuestre que $F(\vec{X})$ es real.

5.3 Diagonalización de una forma cuadrática

Sea $F\big(\vec{X}\big) = \vec{X}^t A\vec{X}$ una forma cuadrática con n variables. Para simplificarla, se pasa de las variables $x_1, x_2, \ldots, x_n$ a las variables $y_1, y_2, \ldots, y_n$ y se supone que las variables anteriores están relacionadas con las nuevas mediante la fórmula $\vec{X} = P\vec{Y}$, donde P es una matriz no singular. Entonces

$$\begin{aligned} \vec{X}^t A\vec{X} &= \big(P\vec{Y}\big)^t A\big(P\vec{Y}\big) && \text{puesto que } \vec{X} = P\vec{Y} \\ &= \big(\vec{Y}^t\ P^t\big)A\big(P\vec{Y}\big) \\ &= \vec{Y}^t\big(P^t AP\big)\vec{Y} = \vec{Y}^t B\vec{Y} && \text{donde } B \text{ es congruente a } A. \end{aligned}$$

Así, $F\big(\vec{X}\big)$ es equivalente a una forma cuadrática $F\big(\vec{Y}\big)$ cuya matriz es B. En las nuevas variables no hay términos mixtos cuando la matriz B sea diagonal. A este proceso se le llama *diagonalización de una forma cuadrática.*

5.3.1 Diagonalización por completación de cuadrados

Un procedimiento para diagonalizar una forma cuadrática es la generalización de la técnica familiar de completar cuadrados, aprendido en el álgebra elemental. El método que se va a estudiar a continuación consiste

en obtener una expresión canónica para $F(\vec{X}) = \vec{X}^t A\vec{X}$ en términos de los menores de la matriz asociada. Para facilitar la comprensión, se comenzará aplicando este método a las formas cuadráticas de dos y tres variables.

Caso I. Si se considera una forma cuadrática en dos variables

$$F(\vec{X}) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2, \tag{5.17}$$

entonces $F(\vec{X})$ se puede expresar como

$$\vec{X}^t A\vec{X} = [x_1, x_2] \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Si cualquiera a_{11} o a_{22} es no nulo, sin pérdida de generalidad se puede asumir que a_{11} es distinto de cero. Entonces, (5.17) se puede escribir como

$$\begin{aligned} F(\vec{X}) &= a_{11}\left[x_1^2 + 2\frac{a_{12}}{a_{11}}x_1x_2 + \left(\frac{a_{12}}{a_{11}}\right)^2 x_2^2 - \left(\frac{a_{12}}{a_{11}}\right)^2 x_2^2 + \frac{a_{22}}{a_{11}}x_2^2\right] \\ &= a_{11}\left\{\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \left[\frac{a_{22}}{a_{11}} - \left(\frac{a_{12}}{a_{11}}\right)^2\right]x_2^2\right\} \\ &= a_{11}\left\{\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \frac{1}{a_{11}^2}\left[a_{11}a_{22} - a_{12}^2\right]x_2^2\right\}. \end{aligned} \tag{5.18}$$

Si se definen

$$y_1 = x_1 + \frac{a_{12}}{a_{11}}x_2 \qquad \text{y} \qquad y_2 = x_2,$$

se tiene que

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = P\vec{X} = \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Entonces, (5.18) en términos de las nuevas variables queda

$$F(\vec{Y}) = a_{11}y_1^2 + \frac{\det A}{a_{11}}\, y_2^2, \tag{5.19}$$

y la forma (5.17) ha sido diagonalizada. La transformación de variables es no singular ($\det P = 1$), pero no es ortogonal. Además, los coeficientes de y_1^2, y_2^2 en (5.19) no son, en general, los valores propios de la matriz asociada a la forma cuadrática A.

El lector puede verificar que cuando $a_{22} \neq 0$, entonces (5.17) se puede escribir como

$$F(\vec{X}) = a_{22}\left\{\left(x_2 + \frac{a_{12}}{a_{22}}x_1\right)^2 + \frac{1}{a_{22}^2}\left[a_{11}a_{22} - a_{12}^2\right]x_1^2\right\}. \quad (5.20)$$

Si se define

$$y_1 = x_2 + \frac{a_{12}}{a_{22}}x_1 \qquad \text{y} \qquad y_2 = x_1,$$

dicha transformación de variables se puede expresar como

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = P\vec{X} = \begin{bmatrix} \frac{a_{12}}{a_{11}} & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Esta transformación de variables es no singular ($\det P = -1$), pero no es ortogonal. Al reemplazar en (5.20), se obtiene

$$F(\vec{Y}) = a_{22}y_1^2 + \frac{\det A}{a_{22}}\, y_2^2. \quad (5.21)$$

En el caso de que a_{11}, a_{22} ambas desaparezcan, el procedimiento anterior no se puede trabajar. Cuando $a_{11} = a_{22} = 0$, la expresión (5.17) se vuelve

$$F(\vec{X}) = 2a_{12}x_1x_2. \quad (5.22)$$

Ahora, se hace la transformación

$$x_1 = y_1 + y_2 \qquad \text{y} \qquad x_2 = y_1 - y_2,$$

la cual se puede expresar matricialmente como

$$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = P\vec{Y} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Esta es una transformación no singular, la cual reduce (5.22) a

$$F(\vec{Y}) = 2a_{12}\left(y_1^2 - y_2^2\right). \quad (5.23)$$

En este caso, también la forma ha sido diagonalizada.

Caso II. Si se considera una forma cuadrática en tres variables

$$F(\vec{X}) = \sum_{i=1}^{3} a_{ii}x_i^2 + 2\sum_{i=1}^{2}\sum_{j>i}^{3} a_{ij}x_ix_j, \tag{5.24}$$

$F(\vec{X})$ se puede expresar como $\vec{X}^tA\vec{X}$, donde

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \qquad \text{y} \qquad \vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Si cualquiera a_{11}, a_{22} o a_{33} es no nulo, sin pérdida de generalidad se puede suponer que $a_{11} \neq 0$. Entonces, (5.24) se puede escribir como

$$\begin{aligned}
F(\vec{X}) &= a_{11}\left[x_1^2 + 2x_1\sum_{k=2}^{3}\frac{a_{1k}}{a_{11}}x_k + \left(\sum_{k=2}^{3}\frac{a_{1k}}{a_{11}}x_k\right)^2 - \left(\sum_{k=2}^{3}\frac{a_{1k}}{a_{11}}x_k\right)^2 \right.\\
&\quad \left. + \sum_{k=2}^{3}\frac{a_{kk}}{a_{11}}x_k^2 + 2\frac{a_{23}}{a_{11}}x_2x_3\right]\\
&= a_{11}\left\{\left(x_1 + \sum_{k=2}^{3}\frac{a_{1k}}{a_{11}}x_k\right)^2 + \left[\sum_{k=2}^{3}\left(\frac{a_{kk}}{a_{11}} - \left(\frac{a_{1k}}{a_{11}}\right)^2\right)x_k^2\right]\right.\\
&\quad \left. + 2\left(\frac{a_{23}}{a_{11}} - \frac{a_{12}a_{13}}{a_{11}^2}\right)x_2x_3\right\}\\
&= a_{11}\left\{\left(x_1 + \sum_{k=2}^{3}\frac{a_{1k}}{a_{11}}x_k\right)^2 + \left[\frac{a_{33}}{a_{11}} - \left(\frac{a_{13}}{a_{11}}\right)^2\right]x_3^2\right.\\
&\quad + \left[\frac{a_{22}}{a_{11}} - \left(\frac{a_{12}}{a_{11}}\right)^2\right]\left[x_2 + \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2}\,x_3\right]^2\\
&\quad \left. - \left[\frac{a_{22}}{a_{11}} - \left(\frac{a_{12}}{a_{11}}\right)^2\right]\left(\frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2}\right)^2 x_3^2\right\}\\
&= a_{11}\left\{\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3\right)^2 + \right.\\
&\quad \left[\frac{a_{22}}{a_{11}} - \left(\frac{a_{12}}{a_{11}}\right)^2\right]\left[x_2 + \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2}\,x_3\right]^2 +\\
&\quad \left.\frac{1}{a_{11}^2}\left[(a_{33}a_{11} - a_{13}^2) - \frac{(a_{11}a_{23} - a_{12}a_{13})^2}{a_{11}a_{22} - a_{12}^2}\right]x_3^2\right\}.
\end{aligned}$$

Ahora, se asume que $M_{33}(A) \neq 0$

$$F(\vec{X}) = a_{11}\left\{\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3\right)^2 + \frac{1}{a_{11}^2}\,\frac{a_{11}\det A}{M_{33}(A)}\,x_3^2 + \frac{1}{a_{11}^2}M_{33}(A)\left[x_2 + \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2}\,x_3\right]^2\right\}. \qquad (5.25)$$

Con la sustitución

$$\begin{aligned} y_1 &= x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3, \\ y_2 &= x_2 + \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2}x_3, \qquad\qquad y_3 = x_3, \end{aligned}$$

se tiene que

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P\vec{X} = \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} \\ 0 & 1 & \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{22} - a_{12}^2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Entonces, (5.25) en términos de las nuevas variables queda

$$F(\vec{Y}) = a_{11}y_1^2 + \frac{M_{33}(A)}{a_{11}}\,y_2^2 + \frac{\det A}{M_{33}(A)}\,y_3^2, \qquad (5.26)$$

y la forma (5.24) ha sido diagonalizada. La transformación de variables es no singular ($\det P = 1$), pero no es ortogonal. Los coeficientes de y_1^2, y_2^2, y_3^2 en (5.26) no son, en general, los valores propios de A.

Se puede verificar fácilmente que cuando $M_{22}(A) \neq 0$, (5.24) se puede escribir como

$$F(\vec{X}) = a_{11}\left\{\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3\right)^2 + \frac{1}{a_{11}}\,\frac{\det A}{M_{22}(A)}\,x_2^2 + \frac{1}{a_{11}^2}M_{22}(A)\left[x_3 + \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{33} - a_{13}^2}\,x_2\right]^2\right\}. \qquad (5.27)$$

Si se definen

$$\begin{aligned} y_1 &= x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3, \\ y_2 &= x_3 + \frac{a_{11}a_{23} - a_{12}a_{13}}{a_{11}a_{33} - a_{13}^2}x_2, \qquad\qquad y_3 = x_2, \end{aligned}$$

dicha transformación de variables se puede expresar como

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P\vec{X} = \begin{bmatrix} 1 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} \\ 0 & \frac{a_{11}a_{23}-a_{12}a_{13}}{a_{11}a_{22}-a_{12}^2} & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Esta transformación de variables es no singular ($\det P = -1$), pero no es ortogonal. Al reemplazar en (5.27), se obtiene

$$F(\vec{Y}) = a_{11}y_1^2 + \frac{M_{22}(A)}{a_{11}}\, y_2^2 + \frac{\det A}{M_{22}(A)}\, y_3^2. \tag{5.28}$$

El procedimiento que se describió puede generalizarse para diagonalizar cualquier forma cuadrática con n variables de la siguiente manera:

Teorema 5.6 Método de reducción de Lagrange

Sea $F(\vec{X})$ una forma cuadrática en $\mathbb{R}^n$ con matriz asociada $A = [a_{ij}]$, luego:

$$F(\vec{X}) = \vec{X}^t A\vec{X} = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} a_{ij}x_i x_j.$$

Entonces, existe una matriz triangular superior T con elementos en la diagonal iguales a 1, tal que el cambio de coordenadas $\vec{X} = T\vec{Y}$ transforma a $\vec{X}^t A\vec{X}$ en:

$$F\big(\vec{Y}\big) = \vec{Y}^t T^t AT\vec{Y} = \Delta_1 y_1^2 + \frac{\Delta_2}{\Delta_1}\, y_2^2 + \frac{\Delta_3}{\Delta_2}\, y_3^2 + \ldots + \frac{\Delta_n}{\Delta_{n-1}}\, y_n^2, \tag{5.29}$$

donde $\Delta_i \neq 0$, $i = 1, 2, \ldots, n$ son los determinantes de las submatrices angulares $A_{[i]}$ de A. ($\Delta_n = \det A$).

Demostración.

La prueba se hace por inducción. Es claro que para $n = 1$, se cumple trivialmente. Para $n = 2$ y $n = 3$ se obtienen las formas cuadráticas (5.19)

y (5.26), respectivamente. Supóngase que es cierto para $n-1$, es decir, existe una matriz de tamaño $\big(n-1\big)\times\big(n-1\big)$,

$$T_{n-1}=\begin{bmatrix}1 & t_{12} & t_{13} & \dots & t_{1n-1}\\ 0 & 1 & t_{23} & \dots & t_{2n-1}\\ 0 & 0 & 1 & \dots & t_{3n-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 1\end{bmatrix},$$

tal que si $\vec{X}^t_{n-1}=(x_1,x_2,\dots,x_{n-1})\in\mathbb{R}^{n-1}$ y $A_{[n-1]}$ es la submatriz angular de A de orden $n-1$, usando el cambio de coordenadas $\vec{X}_{n-1}=T_{n-1}\vec{Y}_{n-1}$, se tiene que

$$\vec{X}^t_{n-1}A_{[n-1]}\vec{X}_{n-1}=\Delta_1y_1^2+\frac{\Delta_2}{\Delta_1}\,y_2^2+\frac{\Delta_3}{\Delta_2}\,y_3^2+\ldots+\frac{\Delta_{n-1}}{\Delta_{n-2}}\,y_{n-1}^2,\quad(5.30)$$

y la matriz A se puede escribir como

$$A=\begin{bmatrix}A_{[n-1]} & \vdots & \vec{U}\\ \dots & . & \dots\\ \vec{U}^t & \vdots & a_{nn}\end{bmatrix},\qquad\text{donde}\qquad\vec{U}=\begin{bmatrix}a_{1n}\\ a_{2n}\\ \vdots\\ a_{n-1n}\end{bmatrix}.$$

Si para cualquier $\vec{c}\in\mathbb{R}^{n-1}$, se considera la matriz triangular

$$T=\begin{bmatrix}T_{n-1} & \vdots & \vec{c}\\ \dots & . & \dots\\ \vec{0}^t & \vdots & 1\end{bmatrix},$$

entonces

$$F(\vec{X})=\vec{X}^tA\vec{X}=\vec{Y}^tT^tAT\vec{Y}=\vec{Y}^t\begin{bmatrix}B_{11} & \vdots & B_{12}\\ \dots & . & \dots\\ B_{12}^t & \vdots & B_{22}\end{bmatrix}\vec{Y},\qquad(5.31)$$

donde

$$\begin{aligned}B_{11}&=T^t_{n-1}A_{[n-1]}T_{n-1}\ \text{ es una matriz de tamaño }\big(n-1\big)\times\big(n-1\big)\\ B_{12}&=T^t_{n-1}A_{[n-1]}\vec{c}+T^t_{n-1}\vec{U}\ \text{ es una matriz de tamaño }\big(n-1\big)\times 1\\ B_{22}&=\vec{c}^{\,t}A_{[n-1]}\vec{c}+2\vec{c}^{\,t}\vec{U}+a_{nn}\ \text{ es un escalar.}\end{aligned}$$

Por la hipótesis de inducción, expresión (5.30), resulta que

$$[y_1,\ldots,y_{n-1}]\,B_{11}\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_{n-1}\end{bmatrix}=[y_1,\ldots,y_{n-1}]\begin{bmatrix}\Delta_1 & 0 & \ldots & 0\\ 0 & \frac{\Delta_2}{\Delta_1} & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \frac{\Delta_{n-1}}{\Delta_{n-2}}\end{bmatrix}\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_{n-1}\end{bmatrix},$$

y como B_{22} es un escalar, para completar la prueba y obtener (5.29), bastará encontrar un $\vec{c}\in\mathbb{R}^n$ para que el vector columna B_{12} sea nulo. Para ello, dado que

$$B_{12}\ =T_{n-1}^t\big(A_{[n-1]}\vec{c}+\vec{U}\big)\qquad\text{y}\qquad T_{n-1}\text{ es no singular,}$$

$\vec{c}$ debe ser tal que

$$A_{[n-1]}\vec{c}+\vec{U}=\vec{0}.$$

Como $\det\big(A_{[n-1]}\big)\neq 0$, existe un único $\vec{c}=-A_{[n-1]}^{-1}\vec{U}$ para el cual B_{12} es un vector columna nulo. Si se reemplaza en B_{22}, se tiene que

$$B_{22}=a_{nn}-\vec{U}^tA_{[n-1]}^{-1}\vec{U},$$

y usando el Teorema 1.21, se tiene que

$$\det\big(A\big)=\det\big(A_{[n-1]}\big)\big(a_{nn}-\vec{U}^tA_{[n-1]}^{-1}\vec{U}\big),$$

luego

$$B_{22}=\frac{\det\big(A\big)}{\det\big(A_{[n-1]}\big)}=\frac{\Delta_n}{\Delta_{n-1}}.$$

Al sustituir en la expresión (5.31), resulta

$$F(\vec{X})=\vec{Y}^tT^tAT\vec{Y}=\Delta_1y_1^2+\frac{\Delta_2}{\Delta_1}\,y_2^2+\frac{\Delta_3}{\Delta_2}\,y_3^2+\ldots+\frac{\Delta_n}{\Delta_{n-1}}\,y_n^2.$$

Esto completa la prueba. ■

Ejemplo 5.7 *Sea* $F(\vec{X})=3x_1^2+3x_2^2+2x_1x_2$. *Encuentre una diagonalización por el método descrito.*

Solución.

Utilizando el resultado obtenido en (5.10), para $n = 2$, se tiene que

$$\vec{X}^t A\vec{X} = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2.$$

Al comparar los a_{ij} con los coeficientes de la función $F(\vec{X})$, se obtiene la matriz

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix},$$

luego $F(\vec{X})$ se puede expresar como

$$\vec{X}^t A\vec{X} = [x_1, x_2] \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{5.32}$$

En este caso, ya que $\Delta_1 = a_{11} = 3 \neq 0$ y $\Delta_2 = \det A = 8$, la forma (5.32) se puede diagonalizar de la siguiente manera:

$$F(\vec{Y}) = 3y_1^2 + \frac{8}{3}y_2^2,$$

donde

$$y_1 = x_1 + \frac{1}{3}x_2 \qquad \text{y} \qquad y_2 = x_2.$$

En forma matricial, se obtiene

$$\vec{Y} = \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & 1 \end{bmatrix} \vec{X} \qquad \text{o} \qquad \vec{X} = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 1 \end{bmatrix} \vec{Y}.$$

Aquí, se usó el hecho de que la transformación tenía determinante 1 y, por lo tanto, era no singular. De este modo

$$\begin{aligned} F(\vec{X}) &= \vec{X}^t \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \vec{X} = \vec{Y}^t \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 1 \end{bmatrix}^t \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 1 \end{bmatrix} \vec{Y} \\ &= \vec{Y}^t \begin{bmatrix} 1 & 0 \\ -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 1 & \frac{8}{3} \end{bmatrix} \vec{Y} = \vec{Y}^t \begin{bmatrix} 3 & 0 \\ 0 & \frac{8}{3} \end{bmatrix} \vec{Y}. \end{aligned}$$

Ejemplo 5.8 *Considere la forma cuadrática dada en el Ejemplo 5.4 y determine una forma diagonal equivalente.*

Solución.

Haciendo referencia al Ejemplo 5.4, la $F(\vec{X})$ se puede escribir como

$$\vec{X}^t A\vec{X} = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{5.33}$$

Dado que

$$\Delta_1 = a_{11} = 2 \neq 0, \quad \Delta_2 = M_{33}(A) = 6 \quad \text{y} \quad \Delta_3 = \det A = 7,$$

se tiene que

$$F(\vec{Y}) = 2y_1^2 + 3y_2^2 + \frac{7}{6}\, y_3^2,$$

donde

$$y_1 = x_1 + x_2 + \frac{1}{2}x_3, \qquad y_2 = x_2 + \frac{1}{3}x_3 \qquad \text{y} \qquad y_3 = x_3.$$

En forma matricial, se tiene que

$$\vec{Y} = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{X} \qquad \text{o} \qquad \vec{X} = \begin{bmatrix} 1 & -1 & -\frac{1}{6} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{Y}.$$

En la transformación anterior, se empleó el hecho de que la transformación tenía determinante 1 y, por lo tanto, era no singular. De este modo

$$\begin{aligned} F(\vec{X}) &= \vec{X}^t A\vec{X} = \vec{Y}^t \begin{bmatrix} 1 & -1 & -\frac{1}{6} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix}^t \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & -\frac{1}{6} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{Y} \\ &= \vec{Y}^t \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -\frac{1}{6} & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 2 & 3 & 0 \\ 1 & 1 & \frac{7}{6} \end{bmatrix} \vec{Y} = \vec{Y}^t \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \frac{7}{6} \end{bmatrix} \vec{Y}. \end{aligned}$$

Teorema 5.7 *Sea $F(\vec{X})$ una forma cuadrática asociada a una matriz simétrica real A. Sea L una matriz no singular triangular inferior tal que A se pueda factorizar como LDL^t. Entonces, el cambio de coordenadas:*

$$\vec{Y} = L^t \vec{X} \tag{5.34}$$

transforma a $\vec{X}^t A\vec{X}$ en $\vec{Y}^t D\vec{Y}$.

Demostración.
La matriz A asociada a la forma se puede factorizar como

$$A = LDU.$$

Como A es simétrica, por el Teorema 3.5, $U = L^t$. Por lo tanto,

$$\begin{aligned}\vec{X}^t A\vec{X} &= \vec{X}^t\big(LDL^t\big)\vec{X} && \text{puesto que } A = LDL^t\\ &= \big(\vec{X}^t L\big)D\big(L^t\vec{X}\big)\\ &= \big(L^t\vec{X}\big)^t D\big(L^t\vec{X}\big) = \vec{Y}^t D\vec{Y}, && \text{puesto que } \vec{Y} = L^t\vec{X}.\end{aligned}$$

Así, queda probado el teorema. ■

A continuación, se presenta una versión de este método de diagonalización:

Procedimiento para diagonalizar una forma cuadrática

i) Halle la matriz de coeficientes simétrica A asociada a $F(\vec{X})$.
ii) Obtenga la descomposición LDL^t de A sin efectuar intercambios de filas que destruyan la simetría, y con elementos en $D = diag\{d_{11}, d_{22}, \ldots, d_{nn}\}$ no necesariamente distintos de cero. Forme L de manera que $\det(L) = 1$.
iii) Transforme a $F(\vec{X})$ en $d_{11}y_1^2 + d_{22}y_2^2 + \ldots + d_{nn}y_n^2$, bajo el cambio de coordenadas $\vec{Y} = L^t\vec{X}$.

Ejemplo 5.9 *Considere la forma cuadrática dada en el Ejemplo 5.7 y determine una forma diagonal equivalente por el método descrito.*

Solución.
La factorización LDL^t de la matriz asociada a la forma cuadrática es

$$\begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix} = \begin{bmatrix}1 & 0\\ \frac{1}{3} & 1\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & \frac{8}{3}\end{bmatrix}\begin{bmatrix}1 & \frac{1}{3}\\ 0 & 1\end{bmatrix}.$$

Luego, la $F(\vec{X})$ se puede expresar como

$$\vec{X}^t A\vec{X} = [x_1, x_2]\begin{bmatrix}1 & 0\\ \frac{1}{3} & 1\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & \frac{8}{3}\end{bmatrix}\begin{bmatrix}1 & \frac{1}{3}\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}. \tag{5.35}$$

Como L es no singular, se hace

$$\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}1 & \frac{1}{3}\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix},$$

y $F(\vec{X})$ se puede escribir en términos de las variables y_1, y_2 como

$$F(\vec{Y}) = 3y_1^2 + \frac{8}{3}y_2^2.$$

Ejemplo 5.10 *Considere la forma cuadrática dada en el Ejemplo 5.4 y determine una forma diagonal equivalente.*

Solución.

La factorización LDL^t de la matriz asociada a la forma cuadrática es

$$\begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \frac{7}{6} \end{bmatrix} \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix},$$

de modo que

$$F(\vec{X}) = \vec{X}^t \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \vec{X} = \vec{X}^t \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \frac{7}{6} \end{bmatrix} \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{X}.$$

Nótese que L es no singular. Por lo tanto, se puede hacer el cambio de variable:

$$\vec{Y} = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \vec{X}$$

y se puede escribir $F(\vec{X})$ en términos de las variables y_1, y_2, y_3 como

$$F(\vec{Y}) = 2y_1^2 + 3y_2^2 + \frac{7}{6}y_3^2.$$

5.3.2 Diagonalización por transformación ortogonal

En la diagonalización de la forma cuadrática $F(\vec{X}) = \vec{X}^t A\vec{X}$ a la forma $F(\vec{Y}) = \vec{Y}^t B\vec{Y}$, por el método de completación de cuadrados, solo se exigió que la matriz P fuera no singular. Puesto que la matriz A asociada a la forma cuadrática es simétrica, entonces se puede diagonalizar a $F(\vec{X})$ mediante una matriz P ortogonal, en cuyo caso los elementos sobre la diagonal de la matriz B son los valores propios de la matriz A.

Teorema 5.8 Teorema de los ejes principales

Sea $F(\vec{X})$ una forma cuadrática asociada a una matriz simétrica real A con valores propios (no necesariamente distintos) $\lambda_1, \lambda_2, \ldots, \lambda_n$. Sea Q una matriz ortogonal propia que diagonaliza a A. Entonces el cambio de coordenadas:

$$\vec{X} = Q\vec{Y} \tag{5.36}$$

transforma a $\vec{X}^t A\vec{X}$ en $\vec{Y}^t D\vec{Y}$, donde

$$D = Q^t AQ = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\}.$$

Demostración.

La demostración consiste en un cálculo directo

$$\begin{aligned} \vec{X}^t A\vec{X} &= \big(Q\vec{Y}\big)^t A\big(Q\vec{Y}\big) && \text{puesto que } \vec{X} = Q\vec{Y} \\ &= \big(\vec{Y}^t\ Q^t\big)A\big(Q\vec{Y}\big) \\ &= \vec{Y}^t\big(Q^t AQ\big)\vec{Y} = \vec{Y}^t D\vec{Y} && \text{puesto que } Q \text{ diagonaliza a } A. \end{aligned}$$

Así, queda el teorema probado. ■

Nota 5.1 Si Q es una matriz ortogonal impropia, es decir, $\det Q = -1$, se deben cambiar de signo todos los elementos de un solo vector columna (o intercambiar dos vectores columnas de Q).

El Teorema 5.8 se llama Teorema de los ejes principales porque define nuevos ejes (los ejes principales) con respecto a los cuales la forma cuadrática tiene una expresión particularmente simple.

A continuación, se presenta una versión de este método de diagonalización:

Procedimiento para diagonalizar una forma cuadrática

i) Halle la matriz de coeficientes simétrica A asociada a $F(\vec{X})$.

ii) Encuentre los valores propios (no necesariamente distintos), $\lambda_1, \lambda_2, \ldots, \lambda_n$ de A.

iii) Obtenga una base ortonormal para $\mathbb{R}^n$ formada por los vectores propios normalizados de A.

iv) Forme la matriz Q cuyas columnas sean los vectores de la base hallada en el paso *iii*), en el orden correspondiente al listado de los valores propios en el paso *ii*). La transformación $\vec{X} = Q\vec{Y}$ es una *rotación* si $\det(Q) = 1$.

v) Transforme a $F\big(\vec{X}\big)$ en $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2$.

Ejemplo 5.11 *Determine los ejes principales de la forma cuadrática dada en el Ejemplo 5.7.*

Solución.

En el Ejemplo 5.7, se obtuvo que la $F\big(\vec{X}\big)$ se puede escribir como

$$\vec{X}^t A \vec{X} = [x_1, x_2] \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{5.37}$$

En este caso, los valores propios de A son $\lambda_1 = 2$ y $\lambda_2 = 4$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, respectivamente.

Para encontrar Q, como $\|\vec{v}_1\| = \sqrt{2}$, se hace $\vec{u}_1 = \dfrac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$, y como $\|\vec{v}_2\| = \sqrt{2}$, se tiene $\vec{u}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. Se puede verificar que la base obtenida para $\mathbb{R}^2$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_2 = 0$. Por lo tanto,

$$Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

Como el $\det(Q) = -1$, intercambiamos las columnas y se tiene

$$Q = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \qquad \text{y} \qquad Q^t A Q = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}.$$

Si se definen los ejes principales como sigue:

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = Q^t \vec{X} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

entonces (5.37) se puede escribir en términos de las nuevas variables $y_1,\ y_2$ como $\vec{Y}^t D\vec{Y}$, o sea,

$$4y_1^2 + 2y_2^2. \tag{5.38}$$

Ejemplo 5.12 *Considere la forma cuadrática dada en el Ejemplo 5.4 y determine sus ejes principales.*

Solución.

Haciendo referencia al Ejemplo 5.4, la $F(\vec{X})$ se puede escribir como

$$\vec{X}^t A\vec{X} = [x_1, x_2, x_3] \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{5.39}$$

Del Ejemplo 3.18, se tiene que

$$Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \quad \text{y} \quad Q^t AQ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}.$$

Por lo tanto, (5.39) se puede escribir en términos de las nuevas variables $y_1,\ y_2,\ y_3$ como $\vec{Y}^t D\vec{Y}$, o sea,

$$y_1^2 + y_2^2 + 7y_3^2, \tag{5.40}$$

donde los ejes principales se obtienen como sigue:

$$\vec{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = Q^t \vec{X} = \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

5.4 Ley de la inercia para formas cuadráticas

Sea $F : \mathbb{R}^n \to \mathbb{R}$ una forma cuadrática. Si $\rho\big(A\big) = r$, entonces toda matriz simétrica que represente a F también tiene rango r. En particular, todas las formas diagonales a las que F sea semejante, mediante una transformación lineal real invertible de variables, tendrán exactamente r coeficientes no nulos. Además, todas las formas diagonales a las que reduzcamos F tienen el mismo número de coeficientes positivos y el mismo número de coeficientes negativos, como se afirma en el resultado obtenido por Sylvester.

Teorema 5.9 Ley de la inercia de Sylvester

Sean A una matriz simétrica de tamaño $n \times n$ y P una matriz no singular del mismo tamaño, entonces:

$$In\big(A\big) = In\big(P^T AP\big).$$

Demostración.

Sea Q una matriz ortogonal tal que

$$Q^t AQ = D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

de forma que $\lambda_1, \lambda_2, \dots, \lambda_i$ son positivos, $\lambda_{i+1}, \lambda_{i+2}, \dots, \lambda_{i+j}$ son negativos y el resto nulos.

Sean $\widehat{A}$ la matriz simétrica $P^t AP$ y W una matriz ortogonal tal que

$$W^t \widehat{A} W = \widehat{D} = \begin{bmatrix} \widehat{\lambda}_1 & 0 & \dots & 0 \\ 0 & \widehat{\lambda}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \widehat{\lambda}_n \end{bmatrix},$$

de forma que $\widehat{\lambda}_1, \widehat{\lambda}_2, \dots, \widehat{\lambda}_{\widehat{i}}$ son positivos, $\widehat{\lambda}_{\widehat{i}+1}, \widehat{\lambda}_{\widehat{i}+2}, \dots, \widehat{\lambda}_{\widehat{i}+\widehat{j}}$ son negativos y el resto nulos.

Se prueba por contradicción que $i = \widehat{i}$. Supongamos que $\widehat{i} > i$.

Sean $\vec{q}_1, \vec{q}_2, \dots, \vec{q}_i$ las primeras i columnas de la matriz Q. Sean $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n$ las filas de la matriz $R = W^t P^{-1}$. Se forma una matriz

B de tamaño $n \times n$, cuyas primeras filas sean las primeras i columnas de Q transpuestas y el resto sean las filas $\widehat{i}+1, \widehat{i}+2, \ldots, n$ de R, es decir,

$$B = \begin{bmatrix} \vec{q}_1{}^t \\ \vec{q}_2{}^t \\ \vdots \\ \vec{q}_i{}^t \\ \vec{r}_{\widehat{i}+1} \\ \vec{r}_{\widehat{i}+2} \\ \vdots \\ \vec{r}_n \end{bmatrix}$$

es una matriz real de tamaño $(i+n-\widehat{i}) \times n$, donde $i+n-\widehat{i} < n$. Por consiguiente, el sistema lineal homogéneo cuya matriz de coeficientes es B tiene solución distinta de la trivial y, por tanto, existe un vector $\vec{u} \neq \vec{0}$, tal que $B\vec{u} = \vec{0}$. En otras palabras, el producto de cualquier fila de la matriz B por el vector $\vec{u}$ es cero, es decir

$$\begin{aligned} &\vec{q}_k{}^t \vec{u} = 0 && \text{para } k = 1, 2, \ldots, i \\ &\vec{r}_k{}^t \vec{u} = 0 && \text{para } k = \widehat{i}+1, \widehat{i}+2, \ldots, n. \end{aligned}$$

Al evaluar $\vec{u}\,{}^t A\vec{u}$, se tiene que

$$\vec{u}\,{}^t A\vec{u} = \vec{u}\,{}^t QDQ^t\vec{u},$$

y al sustituir, se obtiene

$$\begin{aligned} \vec{u}\,{}^t A\vec{u} &= \begin{bmatrix} 0 & \ldots & 0 & v_{i+1} & \ldots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ v_{i+1} \\ \vdots \\ v_n \end{bmatrix} \\ &= \sum_{s=i+1}^{n} \lambda_s v_s^2 < 0, \end{aligned} \tag{5.41}$$

donde, $v_s = \vec{q}_s{}^t \vec{u}$ para $s = i+1, i+2, \ldots, n$. Por otra parte,

$$\vec{u}\,{}^t A\vec{u} = \vec{u}\,{}^t \big(P^t\big)^{-1} \widehat{A} P^{-1}\vec{u} = \vec{u}\,{}^t \big(P^t\big)^{-1} W \widehat{D} W^t P^{-1}\vec{u} = \vec{u}\,{}^t R^t \widehat{D} R\vec{u}.$$

Si se denota $w_k = \vec{r}_k\vec{u}$ para $k = 1, 2, \ldots, \widehat{i}$, se tiene

$$\vec{u}\,{}^t A\vec{u} = \vec{u}\,{}^t R^t \widehat{D} R\vec{u}$$

$$= \begin{bmatrix} w_1 & \ldots & w_{\widehat{i}} & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} \widehat{\lambda}_1 & 0 & \ldots & 0 \\ 0 & \widehat{\lambda}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \widehat{\lambda}_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_{\widehat{i}} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sum_{s=1}^{\widehat{i}} \widehat{\lambda}_s w_s^2 > 0,$$

lo cual contradice (5.41).

Análogamente, se demuestra que la hipótesis $\widehat{i} < i$ conduce a una contradicción. Por lo tanto, se debe tener que $\widehat{i} = i$.

Con $\widehat{j}$ y j se procede de la misma forma. ■

Teorema 5.10 Teorema de Euler

Sea $F(\vec{X})$ una forma cuadrática asociada a una matriz simétrica real A. El valor de $F(\vec{X}) = \vec{X}^t A\vec{X}$ en un vector unitario $\vec{U}$ es

$$F(\vec{U}) = \sum_{j=1}^{n} \lambda_j \cos^2\theta_j, \tag{5.42}$$

donde los λ_j son los valores propios de la matriz simétrica A y los ángulos θ_j son los ángulos entre $\vec{X}$ y los vectores propios ortonormalizados $\vec{q}_j$ correspondientes a los λ_j, respectivamente.

Demostración.

Sean $\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_n$ los vectores propios ortonormalizados de la matriz A y sea $\vec{U}$ un vector unitario arbitrario. Supongamos que θ_j representa el ángulo entre $\vec{U}$ y $\vec{q}_j$, así que

$$\cos\theta_j = \vec{q}_j \cdot \vec{U} = \vec{q}_j{}^t \vec{U}.$$

Si se forma $Q = \begin{bmatrix}\vec{q}_1 & \vec{q}_2 & \dots & \vec{q}_n\end{bmatrix}$ y se obtiene $\vec{Y}$ por la transformación $\vec{U} = Q\vec{Y}$ o $\vec{Y} = Q^t\vec{U}$, entonces,

$$\vec{Y} = Q^t\vec{U} = \begin{bmatrix}\vec{q}_1^{\,t}\\ \vec{q}_2^{\,t}\\ \vdots\\ \vec{q}_j^{\,t}\\ \vdots\\ \vec{q}_n^{\,t}\end{bmatrix}\vec{U} = \begin{bmatrix}\cos\theta_1\\ \cos\theta_2\\ \vdots\\ \cos\theta_j\\ \vdots\\ \cos\theta_n\end{bmatrix}.$$

Por lo tanto,

$$F\big(\vec{U}\big) = \vec{U}^t A\vec{U} = \vec{Y}^t Q^t AQ\vec{Y} = \vec{Y}^t D\vec{Y} = \sum_{j=1}^{n}\lambda_j\cos^2\theta_j,$$

y el teorema queda probado. ■

Ejercicios 5.3

1. Diagonalícense cada una de las siguientes formas cuadráticas por completación de cuadrados y mediante transformación ortogonal:

   a. $F\big(x,y\big) = 4x^2 + 4xy + y^2$. *b.* $F\big(x,y\big) = 9x^2 - 24xy + 16y^2$.

   c. $F\big(x,y\big) = x^2 + 8xy + 16y^2$. *d.* $F\big(x,y\big) = x^2 + 2xy + y^2$

   e. $F\big(x,y,z\big) = 6x^2 + 3y^2 - 2z^2 + 12xy - 18yz$.

   f. $F\big(x,y,z\big) = 7x^2 + 7y^2 + 10z^2 - 2xy - 4xz + 4yz$.

   g. $F\big(x,y,z\big) = 2x^2 + 2y^2 + 5z^2 - 4xy + 2xz - 2yz$.

2. Hallar una condición necesaria y suficiente en a, b y c tal que la forma cuadrática $ax^2 + by^2 + cxy$ se pueda expresar como ku^2.

5.5 Clasificación de las formas cuadráticas

En esta sección se clasifican las formas cuadráticas según sus valores posibles. Una forma cuadrática $F : \mathbb{R}^n \to \mathbb{R}$ es una función de valor real con dominio en $\mathbb{R}^n$. Luego, se pueden distinguir varias clases importantes de formas cuadráticas de acuerdo con los valores que estas asumen para diferentes $\vec{X}$, dichos números reales pueden ser mayores que, menores que o iguales a 0. Obviamente, si el vector es $\vec{X} = \vec{0}$, el valor siempre será 0, por lo tanto no se tendrá en cuenta este vector. Por otra parte, si la matriz A es nula, $F(\vec{X})$ siempre dará el valor cero.

Definición 5.12

Dada $F(\vec{X}) = \vec{X}^t A\vec{X}$ *con* $A \neq \mathrm{O}$ *simétrica, se dice que es*

1. **Definida positiva** *si* $F(\vec{X}) > 0$ *para todo* $\vec{X} \neq \vec{0}$ *en* $\mathbb{R}^n$,

2. **Semidefinida positiva** *si* $F(\vec{X}) \geq 0$ *para todo* $\vec{X}$ *en* $\mathbb{R}^n$,

3. **Definida negativa** *si* $F(\vec{X}) < 0$ *para todo* $\vec{X} \neq \vec{0}$ *en* $\mathbb{R}^n$,

4. **Semidefinida negativa** *si* $F(\vec{X}) \leq 0$ *para todo* $\vec{X}$ *en* $\mathbb{R}^n$,

5. **Indefinida** *si* $F(\vec{X})$ *asume ambos valores positivos y negativos.*

La matriz simétrica asociada A se denomina definida positiva, semidefinida positiva, definida negativa, semidefinida negativa o indefinida según sea la forma cuadrática $F(\vec{X})$ que define.

Teorema 5.11 *Sea A una matriz simétrica de tamaño $n \times n$. Entonces, la forma cuadrática $F(\vec{X}) = \vec{X}^t A\vec{X}$ es:*

1. *Definida positiva si y solo si todos los valores propios de A son positivos.*

2. *Definida negativa si y solo si todos los valores propios son negativos.*

3. *Indefinida si y solo si A tiene valores propios positivos y negativos.*

Demostración.

1. Supongamos que A es definida positiva, sea λ un valor propio de A y $\vec{X}$ el vector propio asociado a λ, entonces

$$\vec{X}^t A\vec{X} = \vec{X}^t \lambda \vec{X} = \lambda \vec{X}^t \vec{X}.$$

Luego, $\lambda = \frac{\vec{X}^t A\vec{X}}{\vec{X}^t \vec{X}}$ es positivo puesto que es el cociente de dos números positivos.

Las otras quedan como ejercicio para el lector. ■

Ejemplo 5.13 *Sea $F(\vec{X}) = 3x^2 + 3y^2 + 2xy$. Determine qué clase de forma cuadrática es.*

Solución.

En el Ejemplo 5.11, se obtuvo que los valores propios de la matriz asociada a la forma $F(\vec{X})$ eran 2 y 4. Por lo tanto, dicha forma cuadrática es definida positiva.

Ejemplo 5.14 *Dada $F(\vec{X}) = 2x^2 + 4xy - y^2$, determine qué clase de forma cuadrática es.*

Solución.

La matriz asociada a la forma cuadrática dada es

$$A = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}.$$

En este caso, los valores propios de A son $\lambda_1 = -2$ y $\lambda_2 = 3$. Por lo tanto, dicha forma cuadrática es indefinida.

Teorema 5.12 *Sea $F : \mathbb{V} \to \mathbb{R}$ una forma cuadrática en el espacio vectorial $\mathbb{V}$ de dimensión finita n. La forma $F(\vec{X})$ se dice que es definida*

positiva si y solo si $A = [a_{ij}]$ la matriz asociada a la forma $F(\vec{X})$ tiene la propiedad de que todos los determinantes de sus submatrices angulares son positivos.

Demostración.

La demostración es por inducción sobre n. Para $n = 1$, la forma cuadrática $F\big(\vec{X}\big)$ está dada por

$$F\big(\vec{X}\big) = [x][a][x] = ax^2,$$

en donde $A = [a]$. El teorema afirma en este caso que la forma $F(\vec{X})$ es definida positiva si y solo si $a > 0$, lo cual es claro.

Por el Teorema 5.8, la matriz $A = [a]$ se transforma en la matriz Q^tAQ, en donde Q es una matriz cuadrada de tamaño 1×1 no singular, esto es, $Q = [q], q \neq 0$. De esta manera,

$$Q^tAQ = [q][a][q] = aq^2.$$

En tal caso, siendo $a > 0$ y $q \neq 0$, se tiene que $aq^2 > 0$, así que la afirmación del teorema no depende de la base considerada en $\mathbb{V}$ para el caso $n = 1$.

Supongamos entonces que el teorema es válido para $n = k - 1$ y veamos si se cumple para $n = k$.

Se demostrará primero que si la forma $F(\vec{X})$ es definida positiva, entonces los determinantes $\Delta_1, \Delta_2, \ldots \Delta_k$ son positivos. Dado que la forma $F(\vec{X})$ se puede escribir según (5.2) como

$$F(\vec{X}) \;=\; \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}x_ix_j = \sum_{i=1}^{k} a_{ii}x_i^2 + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} a_{ij}x_ix_j,$$

esta se puede reescribir de la siguiente manera

$$F(\vec{X}) \;=\; \sum_{i=1}^{k-1} a_{ii}x_i^2 + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} a_{ij}x_ix_j + a_{kk}x_k^2. \tag{5.43}$$

Sea $\mathbb{W}$ un subespacio de $\mathbb{V}$ de dimensión $k-1$ y considérese $F^* : \mathbb{W} \to \mathbb{R}$ una forma cuadrática en $\mathbb{W}$ tal que

$$F^*\big(\vec{X}\big) \;=\; \sum_{i=1}^{k-1}\sum_{j=1}^{k-1} a_{ij}x_ix_j.$$

En efecto, la forma $F^*\big(\vec{X}\big)$ es definida positiva. Supóngase lo contrario, entonces existe un vector $\vec{X}^* \in \mathbb{W}$, digamos $\vec{X}^* = (x_1, x_2, \ldots, x_{k-1})^t$ tal que $F^*(\vec{X}^*) \leq 0$. Ahora, formemos el vector $\vec{X} \in \mathbb{V}$ como sigue $\vec{X} = (x_1, x_2, \ldots, x_{k-1}, 0)^t$ y evaluemos $F(\vec{X})$. Según la expresión (5.43), se obtiene que $F(\vec{X}) \leq 0$, lo cual contradice la hipótesis de que la forma $F(\vec{X})$ es definida positiva.

Por lo tanto, según la hipótesis de inducción, los determinantes de las submatrices angulares de la matriz de la forma $F^*(\vec{X})$ son positivos. Estos son: $\Delta_1, \Delta_2, \ldots \Delta_{k-1}$. Falta probar que Δ_k es también positivo.

Pero por el Teorema 5.8, la matriz A asociada a la forma $F(\vec{X})$, mediante el empleo de una matriz no singular Q, se transforma en la matriz

$$D = Q^t A Q.$$

Al tomar el determinante de las matrices en esta última expresión, se obtiene

$$\det(D) = \det(Q^t A Q) = \det(Q^t)\det(A)\det(Q) = \det(A)\big[\det(Q)\big]^2.$$

Pero como $\det(Q) \neq 0$ y $\det(D) = \prod_{i=1}^{k} \lambda_i$, se tiene que[1]

$$\det(A) = \frac{1}{\big[\det(Q)\big]^2}\det(D) = \frac{1}{[\det(Q)]^2}\prod_{i=1}^{k}\lambda_i > 0.$$

Se ha probado así que $\Delta_k = \det(A) > 0$, como se quería.

Se deja como ejercicio para el lector probar que si los determinantes $\Delta_1, \Delta_2, \ldots \Delta_k$ de alguna matriz (arbitraria pero fija) asociada a la forma cuadrática $F(\vec{X})$ son positivos, entonces la forma $F(\vec{X})$ es definida positiva. ■

Ejemplo 5.15 *Determine si la siguiente forma cuadrática es definida positiva:*

$$F(x, y, z) = 2x^2 + 5y^2 + 2z^2 + 4xy + 2xz + 4yz.$$

[1] Véase el Teorema 5.11.

Solución.
La matriz asociada a la forma cuadrática dada es

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

Luego, los determinantes de las submatrices angulares son

$$\Delta_1 = \det(A_{[1]}) \;=\; \big|2\big| = 2 > 0, \qquad \Delta_2 \;=\; \det\big(A_{[2]}\big) = \begin{vmatrix} 2 & 2 \\ 2 & 5 \end{vmatrix} = 6 > 0 \quad \text{y}$$

$$\Delta_3 = \det(A_{[3]}) = \begin{vmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{vmatrix} = 7 > 0.$$

Como los tres determinantes $\Delta_1, \Delta_2, \Delta_3$ son positivos, se concluye, por el Teorema 5.12, que la forma cuadrática $F(\vec{X})$ es definida positiva.

Corolario 5.12.1 *La forma $F : \mathbb{V} \to \mathbb{R}$ es definida negativa si y solo si $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \ldots$ Es decir, si los determinantes $\Delta_1, \Delta_2, \ldots \Delta_n$ alternan sus signos, comenzando con $\Delta_1 < 0$.*

Demostración.
Es claro que la forma $F : \mathbb{V} \to \mathbb{R}$ dada por

$$F(\vec{X}) = \vec{X}^t A \vec{X}$$

es definida negativa si y solo si la forma $-F : \mathbb{V} \to \mathbb{R}$

$$-F(\vec{X}) = \vec{X}^t(-A)\vec{X}$$

es definida positiva. Según el Teorema 5.12, se debe tener entonces que los determinantes

$$\Delta_1 = \big|-a_{11}\big|, \; \Delta_2 = \begin{vmatrix} -a_{11} & -a_{12} \\ -a_{21} & -a_{22} \end{vmatrix}, \; \Delta_3 = \begin{vmatrix} -a_{11} & -a_{12} & -a_{13} \\ -a_{21} & -a_{22} & -a_{23} \\ -a_{31} & -a_{32} & -a_{33} \end{vmatrix}, \ldots,$$

$\Delta_n = (-1)^n \det(A)$ deben ser positivos. Es decir que

$$\Delta_1 \;=\; \det[-a_{11}] = -\det[a_{11}] > 0$$

$$\Delta_2 = \begin{vmatrix} -a_{11} & -a_{12} \\ -a_{21} & -a_{22} \end{vmatrix} = (-1)^2 \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$$

$$\Delta_3 = \begin{vmatrix} -a_{11} & -a_{12} & -a_{13} \\ -a_{21} & -a_{22} & -a_{23} \\ -a_{31} & -a_{32} & -a_{33} \end{vmatrix} = (-1)^3 \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0,$$

y así sucesivamente, lo que prueba el corolario. ■

Teorema 5.13 *La suma de dos cualesquiera matrices definidas positivas del mismo tamaño es definida positiva. Más generalmente, cualquier combinación lineal no negativa de matrices semidefinidas positivas es semidefinida positiva.*

Demostración.

Sean A y B matrices semidefinidas positivas, sean $\alpha,\ \beta \geq 0$. Si se denota $C = \alpha A + \beta B$ y se calcula

$$\vec{X}^t C\vec{X} = \vec{X}^t(\alpha A + \beta B)\vec{X} = \alpha(\vec{X}^t A\vec{X}) + \beta(\vec{X}^t B\vec{X}) \geq 0, \quad \forall \vec{X} \in \mathbb{C}^n,$$

el caso de más de dos sumandos es tratado en el mismo sentido. Si los coeficientes α y β son positivos y si A y B son matrices definidas positivas y además $\vec{X} \neq \vec{0}$, entonces cada término en la suma es positivo. Por consiguiente, una combinación lineal de matrices definidas positivas es definida positiva. ■

Teorema 5.14 *Si A es una matriz simétrica definida positiva de tamaño $n \times n$, entonces cualquier submatriz principal de A es definida positiva.*

Demostración.

Sea K un subconjunto propio de $\{1, 2, \ldots, n\}$ y denotemos por $A_{(K)}$ la submatriz resultante de eliminar las respectivas filas y columnas de la matriz A, indicadas en K. Entonces $A_{(K)}$ es una submatriz principal de A. Nótese que todas las submatrices se pueden obtener de esta manera; el número $\det[A_{(K)}]$ es un menor de A. Sea $\vec{X} \in \mathbb{C}^n$ un vector no nulo con entradas arbitrarias en las componentes indicadas por K y cero en las otras entradas. Denotando por $\vec{X}_{(K)}$ el vector que se obtiene de eliminar los elementos nulos de $\vec{X}$, obsérvese que

$$\vec{X}_{(K)}^H A_{(K)} \vec{X}_{(K)} = \vec{X}^H A\vec{X} > 0.$$

Puesto que $\vec{X}_{(K)} \neq \vec{0}$ es arbitrario, esto significa que $A_{(K)}$ es definida positiva. ■

Teorema 5.15 *La traza, el determinante y todos los menores principales de una matriz simétrica definida positiva son positivos.*

Demostración.

Sea $A \in M_{n\times n}$ una matriz simétrica definida positiva; luego, por el Teorema 2.20, se sabe que la traza y el determinante son respectivamente la suma y el producto de los valores propios, los cuales por el Teorema 5.11 son todos positivos. La otra parte del teorema se obtiene del Teorema 5.14. ■

Ejemplo 5.16 *Determine si la forma cuadrática dada en el Ejemplo 5.15 es definida positiva verificando si la matriz asociada a la forma cumple las condiciones del teorema anterior.*

Solución.

Según el Ejemplo 5.15, la matriz asociada a la forma cuadrática es

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

La $\operatorname{tr}(A) = 9 > 0$, $\det(A) = 7 > 0$ y los determinantes de algunos menores son

$$\det(M_{33}) = \begin{vmatrix} 2 & 2 \\ 2 & 5 \end{vmatrix} = 6, \qquad \det(M_{22}) = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3.$$

Por lo tanto, la forma cuadrática dada sí es definida positiva.

Teorema 5.16 *Sea $A \in M_{m\times m}$ definida positiva y $P \in M_{m\times n}$, entonces P^tAP es semidefinida positiva. Además, $\rho\big(P^tAP\big) = \rho(P)$, así que P^tAP es definida positiva si y solo si P tiene rango n.*

Demostración.

Es claro que P^tAP es simétrica. Para cualquier $\vec{X} \in \mathbb{C}^n$, se tiene que

$$\vec{X}^t P^t AP\vec{X} = \vec{Y}^t A\vec{Y} \geq 0,$$

donde $\vec{Y} = P\vec{X}$ y la desigualdad se sigue porque A es definida positiva. Así que P^tAP es semidefinida positiva. Obsérvese que $\vec{X}^t P^t AP\vec{X} > 0$ si y solo si $P\vec{X} \neq \vec{0}$, ya que A es definida positiva.

Supongamos que $P\vec{X} = \vec{0}$, entonces por supuesto que $P^tAP\vec{X} = \vec{0}$. Recíprocamente, si $P^tAP\vec{X} = \vec{0}$, entonces $\vec{X}^t P^t AP\vec{X} = 0$, y usando el hecho de que A es definida positiva, se concluye que $P\vec{X} = \vec{0}$. Por lo tanto, $P^tAP\vec{X} = \vec{0}$ si y solo si $P\vec{X} = \vec{0}$, lo cual significa que P^tAP y P tienen el mismo espacio nulo (luego, también tienen el mismo rango). ■

Teorema 5.17 Desigualdad de Fischer

Sea A una matriz cuadrada y definida positiva, particionada como

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

donde A_{11} y A_{22} son submatrices cuadradas, entonces:

$$\det A \leq \det\big(A_{11}\big) \det\big(A_{22}\big).$$

Demostración.

Queda como ejercicio para el lector. ■

Ejercicios 5.4

1. Muestre que las entradas de la diagonal de una matriz simétrica definida positiva son números reales positivos.

2. Muestre que los valores propios, traza, determinante y menores principales de una matriz semidefinida positiva son todos no negativos.

3. Muestre que si $A \in M_{2\times 2}$ es definida positiva, entonces $a_{11}a_{22} > |a_{12}|^2$.

4. Si A es de tamaño $m \times n$ de rango $n < m$, muestre que A^tA es definida positiva y que AA^t es semidefinida positiva.

5. Si A es de tamaño $m \times n$ de rango $k < \text{mín}\{m, n\}$, muestre que A^tA y AA^t son semidefinidas positivas.

5.6 Aplicaciones a la geometría analítica

En esta sección se pretende poner al alcance de los lectores un algoritmo proporcionado por el método de "valores propios y vectores propios" para tratar con más generalidad, agilidad y libertad algunos objetos de la geometría analítica de no fácil manipulación por los métodos tradicionales usados para el estudio de las ecuaciones cuadráticas.

Definición 5.13 Ecuación cuadrática

Una ecuación en las variables x y y de la forma:

$$ax^2 + by^2 + cxy + dx + ey + f = 0, \tag{5.44}$$

donde $a,\ b, \ldots, f$ son números reales con al menos uno de los números $a,\ b,\ c$, distinto de cero, se denomina ecuación cuadrática de segundo grado. Esta se puede escribir en forma matricial como

$$\vec{X}^t A\vec{X} + 2\vec{K}^t\vec{X} + f = 0. \tag{5.45}$$

En esta notación, la expresión $\vec{X}^t A\vec{X}$ es la forma cuadrática asociada y el vector $\vec{K} \in \mathbb{R}^2$ es $\vec{K}^t = \begin{bmatrix} \frac{d}{2} & \frac{e}{2} \end{bmatrix}$.

Definición 5.14 Tipos de ecuación cuadrática

Las curvas representadas por la ecuación cuadrática de segundo grado dada en (5.44) *se pueden clasificar según la posición en la cual estén con respecto a un sistema coordenado cartesiano* $\mathbf{X}$*, así:*

1. *Estándar o canónica, si tiene su centro en el origen.*
2. *Trasladada, si tiene su centro en un punto diferente del origen.*
3. *Rotada, si su posición con respecto al sistema* $\mathbf{X}$ *no es canónica ni tampoco trasladada, pero es posible encontrar un sistema* $\mathbf{Y}$ *con el mismo origen del sistema* $\mathbf{X}$*, y tal que los ejes coordenados de* $\mathbf{Y}$ *forman con los ejes coordenados del sistema* $\mathbf{X}$ *un ángulo agudo* θ*, con respecto al cual la curva está en posición canónica.*

Teorema 5.18 *Si en la ecuación general de segundo grado:*

$$ax^2 + by^2 + cxy + dx + ey + f = 0, \tag{5.46}$$

la matriz A *que representa a la forma cuadrática asociada es no singular, entonces al trasladar el origen de coordenadas al punto* $\vec{X}_0 = -A^{-1}\vec{K}$*, el cambio de coordenadas* $\vec{\tilde{X}} = \vec{X} - \vec{X}_0$ *transforma a* (5.46) *en*

$$a\widetilde{x}^2 + b\widetilde{y}^2 + c\widetilde{x}\widetilde{y} + \widetilde{f} = 0, \tag{5.47}$$

donde $\widetilde{f} = f - \vec{K}^t A^{-1}\vec{K}$ *y el vector* $\vec{K}^t = \begin{bmatrix} \frac{d}{2} & \frac{e}{2} \end{bmatrix}$.

Demostración.

La forma cuadrática dada en (5.46) en forma matricial queda

$$\vec{X}^t A\vec{X} + 2\vec{K}^t\vec{X} + f = 0,$$

donde

$$A = \begin{bmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{bmatrix} \qquad \text{y} \qquad \vec{K}^t = \begin{bmatrix} \frac{d}{2} & \frac{e}{2} \end{bmatrix},$$

al sustituir $\vec{X} = \vec{\widetilde{X}} + \vec{X}_0$, se tiene

$$\vec{\widetilde{X}}^t A \vec{\widetilde{X}} + 2\big[A\vec{X}_0 + \vec{K}\big]^t \vec{\widetilde{X}} + \big[\vec{X}_0^t A \vec{X}_0 + 2\vec{K}^t \vec{X}_0 + f\big] = 0, \tag{5.48}$$

luego, $\vec{X}_0$ será un centro si

$$A\vec{X}_0 + \vec{K} = \vec{0}, \qquad \text{es decir,} \qquad \vec{X}_0 = -A^{-1}\vec{K}. \tag{5.49}$$

Por lo tanto,

$$\underbrace{\vec{X}_0^t A \vec{X}_0 + 2\vec{K}^t \vec{X}_0 + f}_{\widetilde{f}} = \vec{K}^t A^{-1} \vec{K} - 2\vec{K}^t A^{-1}\vec{K} + f$$
$$= f - \vec{K}^t A^{-1} \vec{K}.$$

Si $\vec{\widetilde{X}}^t = [\widetilde{x} \quad \widetilde{y}]$, al reemplazar en (5.48), se obtiene

$$a\widetilde{x}^2 + b\widetilde{y}^2 + c\widetilde{x}\widetilde{y} + \widetilde{f} = 0,$$

la cual representa la traslación de origen y es lo que se quería demostrar. ■

Ejemplo 5.17 *Dada la ecuación cuadrática:*

$$3x^2 + 2xy + 3y^2 - 9 = 0, \tag{5.50}$$

elimine el término cruzado xy utilizando el Teorema 5.8, escriba la ecuación en términos de las nuevas variables e identifique la sección cónica obtenida.

Solución.

En el Ejemplo 5.11, se vio que la forma cuadrática asociada $3x^2 + 2xy + 3y^2$ se puede expresar como

$$4u^2 + 2v^2.$$

Luego, la ecuación cuadrática dada en (5.50), se puede escribir como

$$4u^2 + 2v^2 = 9,$$

la cual es la ecuación de una elipse estándar. Puesto que $\vec{K} = \vec{0}$ entonces su centro es el origen del sistema. Por lo tanto, (5.50) es la ecuación de una elipse estándar rotada. Vea la representación en la Figura 5.1:

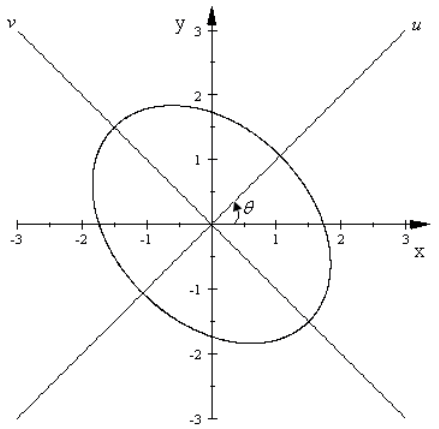

Figura 5.1: Elipse rotada.

Ejemplo 5.18 *Dada la ecuación cuadrática:*

$$2x^2 + 4xy - y^2 - 2x + 3y - 6 = 0, \tag{5.51}$$

elimine el término cruzado xy utilizando el Teorema 5.8, escriba la ecuación en términos de las nuevas variables e identifique la sección cónica obtenida.

Solución.

La forma matricial de esta ecuación cuadrática es

$$\vec{X}^t A \vec{X} + 2\vec{K}^t \vec{X} - 6 = 0, \tag{5.52}$$

donde

$$A = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} \qquad \text{y} \qquad \vec{K}^t = \begin{bmatrix} -1 & \frac{3}{2} \end{bmatrix}.$$

En este caso, los valores propios de A son $\lambda_1 = -2$ y $\lambda_2 = 3$ y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, respectivamente. La identificación de la sección cónica se hace en dos etapas, las cuales se pueden realizar de las siguientes formas:

1. Primero, trasladar el origen de coordenadas al centro de la curva y luego girar los ejes en el ángulo respectivo. Puesto que A es no singular, empleando la expresión (5.49), se determina el centro $\left(\vec{X}_0\right)$ como

$$\vec{X}_0 = -A^{-1}\vec{K} = -\frac{1}{12}\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} -2 \\ 3 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} -2 \\ 5 \end{bmatrix}$$

y $\widetilde{f} = f - \vec{K}^t A^{-1}\vec{K} = -\frac{53}{12}$. Por lo tanto, (5.52) queda

$$\vec{\widetilde{X}}^t A\vec{\widetilde{X}} + \widetilde{f} = \begin{bmatrix} x + \frac{1}{3} & y - \frac{5}{6} \end{bmatrix}\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x + \frac{1}{3} \\ y - \frac{5}{6} \end{bmatrix} - \frac{53}{12} = 0, \quad (5.53)$$

donde $\vec{\widetilde{X}} = \vec{X} - \vec{X}_0$. Para encontrar Q, se usa la expresión (5.7) dada en la definición de inercia, la cual establece primero los valores propios positivos y luego los negativos. De esta manera, como $\|\vec{v}_2\| = \sqrt{5}$, se hace $\vec{u}_2 = \dfrac{\vec{v}_2}{\|\vec{v}_2\|} = \dfrac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ y como $\|\vec{v}_1\| = \sqrt{5}$, se tiene $\vec{u}_1 = \dfrac{1}{\sqrt{5}}\begin{bmatrix} -1 \\ 2 \end{bmatrix}$. Se puede verificar que la base obtenida para $\mathbb{R}^2$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_2 = 0$. Por lo tanto,

$$Q = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \qquad \text{y} \qquad Q^t AQ = D = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}.$$

Dado que el $\det(Q) = 1$ se define

$$\begin{aligned} \vec{\widetilde{X}} = \begin{bmatrix} \widetilde{x} \\ \widetilde{y} \end{bmatrix} &= Q^t\left(\vec{X} - \vec{X}_0\right) = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} x + \frac{1}{3} \\ y - \frac{5}{6} \end{bmatrix} \\ &= \vec{Y} - \vec{Y}_0 = \begin{bmatrix} u & v \end{bmatrix}^t - \begin{bmatrix} \frac{1}{6\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}^t, \end{aligned}$$

donde $\vec{Y} = Q^t\vec{X}$ y $\vec{Y}_0 = Q^t\vec{X}_0$, entonces, la expresión (5.53) se puede escribir en términos de las nuevas variables $\widetilde{x}$, $\widetilde{y}$ como $\vec{\widetilde{X}}^t D\vec{\widetilde{X}}$, es decir,

$$\begin{aligned} 3\widetilde{x}^2 \quad - \quad 2\widetilde{y}^2 \quad &= \frac{53}{12} \\ 3\left(u - \frac{1}{6\sqrt{5}}\right)^2 - 2\left(v - \frac{2}{\sqrt{5}}\right)^2 &= \frac{53}{12}, \end{aligned} \tag{5.54}$$

la cual es la ecuación de una hipérbola con centro en $\left(\frac{1}{6\sqrt{5}}, \frac{2}{\sqrt{5}}\right)$. Por lo tanto, la ecuación (5.51) es una hipérbola rotada y trasladada.

2. Girar los ejes, dándoles las direcciones principales de la curva y después efectuar una traslación de los ejes girados. Este método se puede emplear cuando A es singular, y para ello, se define

$$\vec{Y} = \begin{bmatrix} u \\ v \end{bmatrix} = Q^t\vec{X} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{5.55}$$

Al sustituir en (5.52) el cambio de variable propuesto en (5.55), se tiene

$$2\vec{K}^t\vec{X} \ = 2\vec{K}^tQ\vec{Y} = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 & 3 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 & 8 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}.$$

Luego, la ecuación cuadrática dada en (5.51) se puede reescribir como

$$3u^2 - 2v^2 - \frac{1}{\sqrt{5}}u + \frac{8}{\sqrt{5}}v = 6.$$

Naturalmente, esta no es la ecuación de una cónica en posición canónica, pero sí está trasladada porque al completar los cuadrados, se obtiene

$$\begin{aligned} 3\left[u^2 - \frac{1}{3\sqrt{5}}u + \frac{1}{180}\right] - 2\left[v^2 - \frac{4}{\sqrt{5}}v + \frac{4}{5}\right] &= 6 - \frac{8}{5} + \frac{1}{60} \\ 3\left[u - \frac{1}{6\sqrt{5}}\right]^2 - 2\left[v - \frac{2}{\sqrt{5}}\right]^2 &= \frac{53}{12}. \end{aligned}$$

Nótese que esta cónica es la misma que se obtuvo en (5.54). La representación gráfica aparece en la Figura 5.2:

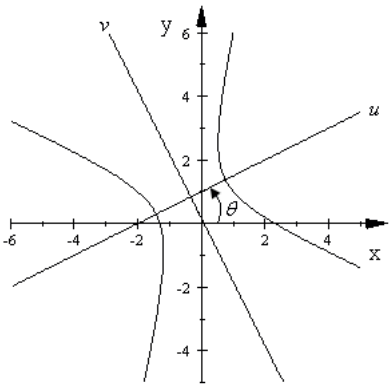

Figura 5.2: Hipérbola rotada y trasladada.

5.6.1 Rotación de ejes en $\mathbb{R}^2$

Como ya se ha señalado, una transformación $\vec{Y} = Q^t\vec{X}$, donde Q es ortogonal, se llama *transformación ortogonal*. Ahora, se examina la interpretación geométrica de estas transformaciones en $\mathbb{R}^2$.

Teorema 5.19 Rotación de ejes en $\mathbb{R}^2$

Sean $\mathfrak{B} = \{\vec{e}_1, \vec{e}_2\}$ *la base del sistema coordenado* $\mathbf{X}$ *y* $\mathfrak{B}' = \{\vec{e}_1', \vec{e}_2'\}$ *la base correspondiente al sistema* $\mathbf{Y}$. *Entonces si*

$$\vec{e}_j' = a_{1j}\vec{e}_1 + a_{2j}\vec{e}_2, \qquad \textit{para} \qquad j = 1, 2,$$

las coordenadas (x_1, x_2) *de un punto cualquiera* P *en el sistema* $\mathbf{X}$ *y las coordenadas* (y_1, y_2) *del mismo punto en el sistema* $\mathbf{Y}$ *están relacionadas como sigue:*

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = A \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

donde $A = [a_{ij}]$ es la matriz de transición (o matriz de cambio de base) de $\mathfrak{B}$ a $\mathfrak{B}'$.

Demostración.

Supongamos que los sistemas coordenados $\mathbf{X}$ y $\mathbf{Y}$ (en $\mathbb{R}^2$) tienen el mismo origen O. Sea $\overrightarrow{OP}$ el vector formado desde el origen del sistema coordenado $\mathbf{X}$ al punto $P = (x_1, x_2)$, entonces

$$\overrightarrow{OP} = x_1\vec{e}_1 + x_2\vec{e}_2.$$

Por otra parte, el vector $\overrightarrow{OP}$ formado desde el origen del sistema coordenado $\mathbf{Y}$ al punto $P = (y_1, y_2)$, es

$$\begin{aligned}\overrightarrow{OP} &= y_1\vec{e}_1' + y_2\vec{e}_2' \\ &= y_1\big(a_{11}\vec{e}_1 + a_{21}\vec{e}_2\big) + y_2\big(a_{12}\vec{e}_1 + a_{22}\vec{e}_2\big) \\ &= \underbrace{\big(y_1a_{11} + y_2a_{12}\big)}\,\vec{e}_1 + \underbrace{\big(y_1a_{21} + y_2a_{22}\big)}\,\vec{e}_2 \\ &= \qquad x_1 \qquad \vec{e}_1 + \qquad x_2 \qquad \vec{e}_2.\end{aligned}$$

Aquí, se usó el hecho de que la representación de $\overrightarrow{OP}$ como combinación lineal de $\vec{e}_1$ y $\vec{e}_2$ es única. Luego,

$$x_1 = a_{11}y_1 + a_{12}y_2 \qquad \text{y} \qquad x_2 = a_{21}y_1 + a_{22}y_2,$$

y el teorema queda probado. ■

5.6.1.1 Cambio de dirección de ejes en $\mathbb{R}^2$ conservando el mismo origen

Consideremos que el sistema coordenado $\mathbf{X}$ (en $\mathbb{R}^2$) es rectangular y tiene unidades iguales sobre ambos ejes. Esto significa que si $\vec{e}_1$ y $\vec{e}_2$ son los vectores base (unitarios), entonces son perpendiculares entre sí.

Supongamos que los ejes coordenados $\mathbf{Y}$ se obtienen haciendo girar el sistema $\mathbf{X}$ un ángulo θ alrededor del origen, en sentido contrario al de las manecillas del reloj, conservando la ortogonalidad. Los vectores base $\vec{e}_1'$ y $\vec{e}_2'$ del sistema $\mathbf{Y}$, forman también una base ortonormal y están dados por

$$\begin{aligned}\vec{e}_1' &= [\cos\theta, \operatorname{sen}\theta] = \cos\theta\vec{e}_1 + \operatorname{sen}\theta\vec{e}_2. \\ \vec{e}_2' &= [-\operatorname{sen}\theta, \cos\theta] = -\operatorname{sen}\theta\vec{e}_1 + \cos\theta\vec{e}_2.\end{aligned}$$

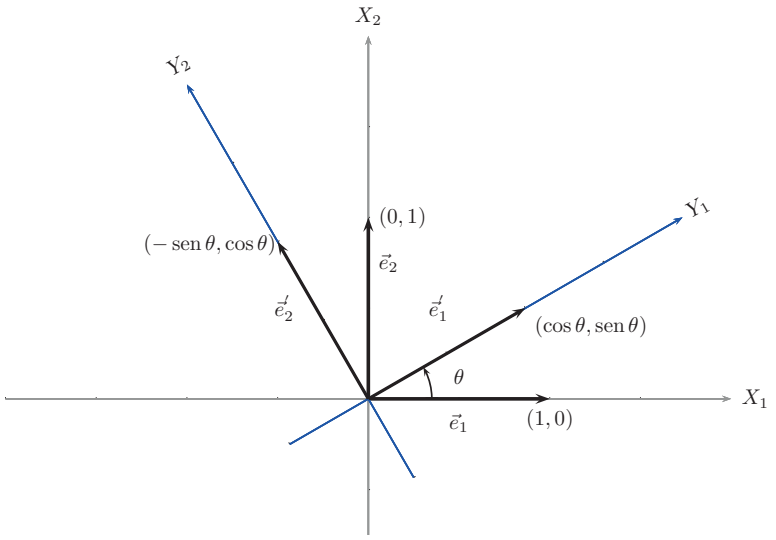

Figura 5.3: Rotación en $\mathbb{R}^2$.

Se deduce por el Teorema 5.19 que las coordenadas de un punto en ambos sistemas están relacionadas por

$$x_1 = y_1 \cos\theta - y_2 \operatorname{sen}\theta \qquad \text{y} \qquad x_2 = y_1 \operatorname{sen}\theta + y_2 \cos\theta, \tag{5.56}$$

que son las ecuaciones de una rotación levógira de ejes, cuando el ángulo girado es θ. Si se denota por A_θ la matriz ortogonal

$$A_\theta = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}, \tag{5.57}$$

entonces (5.56) se puede expresar matricialmente como

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A_\theta \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

la cual es una transformación ortogonal propia, puesto que el $\det A_\theta = 1$ y representa cualquier rotación de ejes en $\mathbb{R}^2$.

A continuación, se presenta un teorema para obtener el ángulo de rotación θ de una ecuación cuadrática de segundo grado.

Teorema 5.20 *Sea la ecuación general de segundo grado:*

$$ax^2 + by^2 + cxy + dx + ey + f = 0. \tag{5.58}$$

Entonces existe un número único θ, llamado ángulo de rotación, tal que

$$\theta = \frac{1}{2}\ \arctan\left(\frac{c}{a-b}\right), \qquad \theta \in \left[0, \frac{\pi}{2}\right). \tag{5.59}$$

Demostración.

La matriz que representa la forma cuadrática dada en (5.58) es

$$A = \begin{bmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{bmatrix}. \tag{5.60}$$

Supongamos que λ es un valor propio de A con vector propio unitario correspondiente $\vec{X}^t = \begin{bmatrix}\cos\theta & \operatorname{sen}\theta\end{bmatrix}$, luego $A\vec{X} = \lambda\vec{X}$. Si se premultiplica por $\vec{X}^t$, se obtiene

$$\vec{X}^t A\vec{X} = \lambda\vec{X}^t\vec{X} = \lambda.$$

Al sustituir $\vec{X}$ y A, se tiene

$$\begin{aligned}\lambda\ &= \begin{bmatrix}\cos\theta & \operatorname{sen}\theta\end{bmatrix}\begin{bmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{bmatrix}\begin{bmatrix}\cos\theta \\ \operatorname{sen}\theta\end{bmatrix} \\ &= a\ \cos^2\theta + c\ \operatorname{sen}\theta\cos\theta + b\operatorname{sen}^2\theta \\ &= (a-b)\ \cos^2\theta + c\ \operatorname{sen}\theta\cos\theta + b.\end{aligned}$$

Si se multiplica por 2 y se usan las siguientes identidades trigonométricas

$$2\cos^2\theta\ =\ \cos(2\theta) + 1 \qquad \text{y} \qquad \operatorname{sen}(2\theta)\ = 2\operatorname{sen}\theta\cos\theta,$$

se llega a

$$2\lambda = (a-b)\ \cos(2\theta) + c\ \operatorname{sen}(2\theta) + (a+b).$$

Pero como λ es valor propio de A, se tiene que

$$2\lambda = \operatorname{tr}(A) \pm \sqrt{\operatorname{tr}^2(A) - 4\det(A)}.$$

Al igualar estas dos expresiones, se obtiene

$$(a-b)\cos(2\theta) + c\operatorname{sen}(2\theta) + (a+b)\ =\ \operatorname{tr}(A) \pm \sqrt{\operatorname{tr}^2(A) - 4\det(A)},$$

pero $\operatorname{tr}(A) = a + b$, por consiguiente

$$\begin{aligned}(a-b)\cos(2\theta) + c\,\operatorname{sen}(2\theta) &= \pm\sqrt{\operatorname{tr}^2(A) - 4\det(A)} \\ \underbrace{\pm\frac{a-b}{\sqrt{\operatorname{tr}^2(A) - 4\det(A)}}}_{u}\cos(2\theta) + &\underbrace{\frac{\pm c}{\sqrt{\operatorname{tr}^2(A) - 4\det(A)}}}_{v}\operatorname{sen}(2\theta) = 1 \\ u\qquad\qquad \cos(2\theta) + &\qquad v\qquad\qquad \operatorname{sen}(2\theta) = 1.\end{aligned}$$

Luego,

$$\begin{aligned}(v\,\operatorname{sen}(2\theta))^2 &= (1 - u\cos(2\theta))^2 \\ v^2\operatorname{sen}^2(2\theta) &= 1 - 2u\cos(2\theta) + u^2\cos^2(2\theta) \\ v^2(1 - \cos^2(2\theta)) &= 1 - 2u\ \cos(2\theta) + u^2\cos^2(2\theta).\end{aligned}$$

El lector puede probar fácilmente que $u^2 + v^2 = 1$, de manera que

$$\begin{aligned}\cos^2(2\theta) - 2u\ \cos(2\theta) + (1 - v^2) &= 0 \\ [\cos(2\theta) - u]^2 &= 0 \\ \therefore\ u &= \cos(2\theta).\end{aligned}$$

Por lo tanto, $v = \operatorname{sen}(2\theta)$ y entonces

$$\frac{v}{u} = \tan(2\theta) = \frac{c}{a-b}. \tag{5.61}$$

Al tomar arctan a ambos lados de (5.61), se obtiene la ecuación (5.59). ■

Nota 5.2 Si la matriz que representa una forma cuadrática tiene la forma (5.60), entonces el ángulo θ de rotación satisface que

$$\left.\begin{aligned}\operatorname{sen}\big(2\theta\big) &= \frac{c}{\sqrt{\operatorname{tr}^2 A - 4\det A}} = \frac{c}{\lambda_1 - \lambda_2} \\ \cos\big(2\theta\big) &= \frac{a-b}{\sqrt{\operatorname{tr}^2 A - 4\det A}} = \frac{a-b}{\lambda_1 - \lambda_2}\end{aligned}\right\}\ \tan\big(2\theta\big) = \frac{c}{a-b}. \tag{5.62}$$

Por otra parte, si se usa la siguiente identidad trigonométrica

$$\tan(2\theta) = \frac{2\tan\theta}{1 - \tan^2\theta},$$

al sustituir en la última expresión dada en (5.62) se tiene que

$$\begin{aligned} \frac{2\tan\theta}{1-\tan^2\theta} &= \frac{c}{a-b} \\ 2(a-b)\tan\theta &= c\big(1-\tan^2\theta\big). \end{aligned}$$

Al resolver esta ecuación cuadrática, se llega a

$$\tan\theta = \frac{b-a}{c} \pm \frac{\sqrt{\operatorname{tr}^2 A - 4\det A}}{c}. \tag{5.63}$$

Ejemplo 5.19 *Determine el ángulo de rotación de la ecuación cuadrática dada en el Ejemplo 5.17.*

Solución.

La matriz simétrica asociada a la forma cuadrática es

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

La ecuación (5.61) no es aplicable ya que como $a = b$ se dividiría por 0, se usa (5.63) y se tiene que

$$\tan\theta = 1. \tag{5.64}$$

Cualquier solución de (5.64) sirve a nuestro propósito. Si se escoge la solución para la cual $0 < \theta < 90°$, entonces $\operatorname{sen}(\theta) = \frac{1}{\sqrt{2}}$ y $\cos(\theta) = \frac{1}{\sqrt{2}}$, es decir, el ángulo de rotación θ vale aproximadamente $\theta = 45°$ y, construyendo la matriz de rotación (5.57), se tiene que

$$A_\theta = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix},$$

la cual coincide con la matriz ortogonal dada en el Ejemplo 5.11.

Ejemplo 5.20 *Determine el ángulo de rotación de la ecuación cuadrática dada en el Ejemplo 5.18.*

Solución.

La matriz simétrica asociada a la forma cuadrática es

$$A = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}.$$

Por la ecuación (5.61), se tiene que

$$\tan(2\theta) = \frac{4}{2-(-1)} = \frac{4}{3}. \tag{5.65}$$

Cualquier solución de (5.65) sirve a nuestro propósito. Si se escoge la solución para la cual $0 < 2\theta < 90°$, entonces $\operatorname{sen}(2\theta) = \frac{4}{5}$ y $\cos(2\theta) = \frac{3}{5}$ y como θ es un ángulo agudo,

$$\operatorname{sen}\theta = \sqrt{\frac{1-\cos 2\theta}{2}} = \frac{1}{\sqrt{5}} \quad \text{y} \quad \cos\theta = \sqrt{\frac{1+\cos 2\theta}{2}} = \frac{2}{\sqrt{5}}.$$

Es decir, el ángulo de rotación θ vale aproximadamente $\theta = 26°33'54.18''$ y, construyendo la matriz de rotación (5.57), se tiene que

$$A_\theta = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix},$$

la cual coincide con la matriz ortogonal dada en el Ejemplo 5.18.

5.6.2 Clasificación de las ecuaciones cuadráticas

Para la ecuación general de segundo grado:

$$ax^2 + by^2 + cxy + dx + ey + f = 0, \tag{5.66}$$

se definen las cantidades

$$A = \begin{bmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{bmatrix}, \qquad \begin{aligned} \omega &= \operatorname{tr}(A), \\ \mu &= \det A \end{aligned} \qquad \text{y} \qquad \nu = \det \begin{bmatrix} A & \vdots & \vec{K} \\ \cdots & \cdot & \cdots \\ \vec{K}^t & \vdots & f \end{bmatrix},$$

donde $\vec{K}^t = \begin{bmatrix} \frac{d}{2} & \frac{e}{2} \end{bmatrix}$ y ω, μ son los coeficientes de $P_A(\lambda)$, entonces la ecuación (5.66) representa los siguientes lugares geométricos (L.G.):

$$\textbf{L.G.} \begin{cases} \mu \neq 0 \begin{cases} \mu > 0 \begin{cases} \nu \neq 0 \begin{cases} \omega\nu < 0 & \text{Una elipse} \\ \omega\nu > 0 & \text{Ninguno} \end{cases} \\ \nu = 0 \quad \text{Un punto} \end{cases} \\ \mu < 0 \begin{cases} \nu \neq 0 & \text{Una hipérbola} \\ \nu = 0 & \text{Dos rectas que se cortan} \end{cases} \end{cases} \\ \mu = 0 \begin{cases} \nu \neq 0 & \text{Una parábola} \\ \nu = 0 & \text{Dos rectas paralelas, o una recta, o ninguno.} \end{cases} \end{cases}$$

El cuadro anterior se acostumbra a interpretar como sigue:

$$\textbf{L.G.} \begin{cases} \textbf{Curva con centro} \ (\mu \neq 0) \begin{cases} \textbf{De tipo elíptico} \ (\mu > 0) \begin{cases} \text{Una elipse } (\omega\nu < 0) \\ \text{Caso degenerado } (\omega\nu \geq 0) \end{cases} \\ \textbf{De tipo hiperbólico} \ (\mu < 0) \begin{cases} \text{Una hipérbola } (\nu \neq 0) \\ \text{Caso degenerado } (\nu = 0) \end{cases} \end{cases} \\ \textbf{Curva sin centro} \ (\mu = 0) \begin{cases} \text{Una parábola } (\nu \neq 0) \\ \text{Caso degenerado } (\nu = 0). \end{cases} \end{cases}$$

Cuando $\nu \neq 0$, la ecuación cuadrática (5.66) también puede ser clasificada de acuerdo con la inercia de la matriz A asociada a la forma cuadrática, como sigue:

Identificación de las ecuaciones cuadráticas	
Inercia	**Nombre de la cónica**
$In(A) = (2, 0, 0)$	Elipse
$In(A) = (1, 1, 0)$	Hipérbola
$In(A) = (1, 0, 1)$	Parábola

Las formas cuadráticas también pueden usarse para analizar ecuaciones de superficies cuádricas en el espacio.

Definición 5.15 Superficie cuádrica

Una ecuación de segundo grado en x, y, y z de la forma:

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0, \qquad (5.67)$$

donde $a,\ b, \ldots, j$ *son números reales y* $|a|+|b|+|c|+|d|+|e|+|f| \neq 0$*, se denomina superficie cuádrica. Esta se puede escribir en forma matricial como*

$$\vec{X}^t A\vec{X} + 2\vec{K}^t\vec{X} + j = 0. \tag{5.68}$$

En esta notación, la expresión $\vec{X}^t A\vec{X}$ *es la forma cuadrática asociada y el vector* $\vec{K} \in \mathbb{R}^3$ *es dado por* $\vec{K}^t = \begin{bmatrix} \frac{g}{2} & \frac{h}{2} & \frac{i}{2} \end{bmatrix}$.

Teorema 5.21 *Si en la superficie cuádrica:*

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0, \tag{5.69}$$

la matriz A *que representa a la forma cuadrática asociada es no singular, entonces al trasladar el origen de coordenadas al punto* $\vec{X}_0 = -A^{-1}\vec{K}$*, el cambio de coordenadas* $\vec{\tilde{X}} = \vec{X} - \vec{X}_0$ *transforma a* (5.69) *en*

$$a\widetilde{x}^2 + b\widetilde{y}^2 + c\widetilde{z}^2 + d\widetilde{x}\widetilde{y} + e\widetilde{x}\widetilde{z} + f\widetilde{y}\widetilde{z} + \widetilde{j} = 0, \tag{5.70}$$

donde $\widetilde{j} = j - \vec{K}^t A^{-1}\vec{K}$ *y el vector* $\vec{K}^t = \begin{bmatrix} \frac{g}{2} & \frac{h}{2} & \frac{i}{2} \end{bmatrix}$.

Demostración.

La superficie cuádrica dada en (5.69) en forma matricial queda

$$\vec{X}^t A\vec{X} + 2\vec{K}^t\vec{X} + j = 0,$$

donde

$$A = \begin{bmatrix} a & \frac{d}{2} & \frac{e}{2} \\ \frac{d}{2} & b & \frac{f}{2} \\ \frac{e}{2} & \frac{f}{2} & c \end{bmatrix} \qquad \text{y} \qquad \vec{K}^t = \begin{bmatrix} \frac{g}{2} & \frac{h}{2} & \frac{i}{2} \end{bmatrix},$$

y al sustituir $\vec{X} = \vec{\tilde{X}} + \vec{X}_0$, se tiene

$$\vec{\tilde{X}}^t A\vec{\tilde{X}} + 2\big[A\vec{X}_0 + \vec{K}\big]^t\vec{\tilde{X}} + \big[\vec{X}_0^t A\vec{X}_0 + 2\vec{K}^t\vec{X}_0 + j\big] = 0. \tag{5.71}$$

Luego, $\vec{X}_0$ será un centro si

$$A\vec{X}_0 + \vec{K} = \vec{0}, \qquad \text{es decir,} \qquad \vec{X}_0 = -A^{-1}\vec{K}. \tag{5.72}$$

Por lo tanto,

$$\begin{aligned} \underbrace{\vec{X}_0^t A\vec{X}_0 + 2\vec{K}^t\vec{X}_0 + j}_{\widetilde{j}} &= \vec{K}^t A^{-1}\vec{K} - 2\vec{K}^t A^{-1}\vec{K} + j \\ &= j - \vec{K}^t A^{-1}\vec{K}. \end{aligned}$$

Si $\vec{\widetilde{X}}^t = [\widetilde{x} \quad \widetilde{y} \quad \widetilde{z}]$, al reemplazar en (5.71), se obtiene

$$a\widetilde{x}^2 + b\widetilde{y}^2 + c\widetilde{z}^2 + d\widetilde{x}\widetilde{y} + e\widetilde{x}\widetilde{z} + f\widetilde{y}\widetilde{z} + \widetilde{j} = 0,$$

la cual representa la traslación de origen y es lo que se quería demostrar. ■

Ejemplo 5.21 *Considere la ecuación cuadrática:*

$$2x^2 + 5y^2 + 2z^2 + 4xy + 2xz + 4yz - 36 = 0. \tag{5.73}$$

Determine la superficie cuádrica obtenida al eliminar los términos de productos cruzados.

Solución.

Haciendo referencia al Ejemplo 5.12, dicha ecuación cuadrática se puede escribir como

$$\vec{X}^t A\vec{X} = 36, \tag{5.74}$$

donde

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

En este caso, (5.73) se puede escribir en términos de las nuevas variables $u,\ v,\ w$ como $\vec{Y}^t D\vec{Y} = 36$, o sea,

$$u^2 + v^2 + 7w^2 = 36. \tag{5.75}$$

En $\mathbb{R}^3$ la superficie definida por (5.75) se llama elipsoide.

Ejemplo 5.22 *Considere la ecuación cuadrática*

$$7x^2 + 7y^2 + 10z^2 - 2xy - 4xz + 4yz - 12x + 12y + 60z - 24 = 0. \quad (5.76)$$

Determine la superficie cuádrica obtenida al eliminar los términos de productos cruzados.

Solución.

La forma matricial de esta ecuación cuadrática es

$$\vec{X}^t A\vec{X} + 2\vec{K}^t\vec{X} - 24 = 0, \quad (5.77)$$

donde

$$A = \begin{bmatrix} 7 & -1 & -2 \\ -1 & 7 & 2 \\ -2 & 2 & 10 \end{bmatrix} \qquad \text{y} \qquad \vec{K}^t = [-6 \quad 6 \quad 30].$$

En este caso, los valores propios de A son $\lambda_1 = 12$ y $\lambda_2 = 6$ (de multiplicidad algebraica 2). El vector propio correspondiente a $\lambda_1 = 12$ es $\vec{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$ y los correspondientes a $\lambda_2 = 6$ son $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$, respectivamente. Para determinar la sección cónica, se hacen en dos etapas, las cuales se pueden realizar de las siguientes formas:

1. Primero, trasladar el origen de coordenadas al centro de la curva y luego girar los ejes en el ángulo respectivo. Puesto que A es no singular, empleando la expresión (5.72), se determina el centro $\left(\vec{X}_0\right)$ como

$$\vec{X}_0 = -A^{-1}\vec{K} = -\frac{1}{72}\begin{bmatrix} 11 & 1 & 2 \\ 1 & 11 & -2 \\ 2 & -2 & 8 \end{bmatrix}\begin{bmatrix} -6 \\ 6 \\ 30 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix}$$

y $\widetilde{j} = j - \vec{K}^t A^{-1}\vec{K} = -114$, por lo tanto la ecuación (5.77) queda

$$\vec{\widetilde{X}}^t A\vec{\widetilde{X}} + \widetilde{j} = \begin{bmatrix} x & y & z+3 \end{bmatrix}\begin{bmatrix} 7 & -1 & -2 \\ -1 & 7 & 2 \\ -2 & 2 & 10 \end{bmatrix}\begin{bmatrix} x \\ y \\ z+3 \end{bmatrix} - 114 = 0, \quad (5.78)$$

donde $\vec{\tilde{X}} = \vec{X} - \vec{X}_0$. Para encontrar la matriz Q, se ortonormalizan los vectores propios de A, como $\|\vec{v}_1\| = \sqrt{6}$ se hace $\vec{u}_1 = \dfrac{\vec{v}_1}{\|\vec{v}_1\|} = \dfrac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$. Después, se aplica el proceso de ortonormalización de Gram-Schmidt a $\{\vec{v}_2, \vec{v}_3\}$ para obtener una base ortonormal del espacio propio $\mathcal{B}_2$. Puesto que $\|\vec{v}_2\| = \sqrt{2}$, se tiene $\vec{u}_2 = \dfrac{\vec{v}_2}{\|\vec{v}_2\|} = \dfrac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Por último,

$$\begin{aligned} \vec{v}_3' &= \vec{v}_3 - (\vec{v}_3 \cdot \vec{u}_2)\vec{u}_2 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Entonces, $\|\vec{v}_3'\| = \sqrt{3}$ y $\vec{u}_3 = \dfrac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. Se puede verificar que la nueva base de $\mathcal{B}_2$ es ortonormal observando que $\vec{u}_2 \cdot \vec{u}_3 = 0$. También, se puede verificar que la base obtenida para $\mathbb{R}^3$ es ortonormal observando que $\vec{u}_1 \cdot \vec{u}_3 = 0$ y $\vec{u}_2 \cdot \vec{u}_3 = 0$. Por tanto,

$$Q = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & \sqrt{3} & \sqrt{2} \\ 1 & \sqrt{3} & -\sqrt{2} \\ 2 & 0 & \sqrt{2} \end{bmatrix}.$$

Como $\det(Q) = -1$, se multiplica la tercera columna por -1 y se tiene

$$Q = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & \sqrt{3} & -\sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \\ 2 & 0 & -\sqrt{2} \end{bmatrix} \quad \text{y} \quad Q^t A Q = D = \begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

Si se define

$$\begin{aligned} \vec{\tilde{X}} = \begin{bmatrix} \widetilde{x} \\ \widetilde{y} \\ \widetilde{z} \end{bmatrix} &= Q^t\big(\vec{X} - \vec{X}_0\big) = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & 1 & 2 \\ \sqrt{3} & \sqrt{3} & 0 \\ -\sqrt{2} & \sqrt{2} & -\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \\ z+3 \end{bmatrix} \\ &= \vec{Y} - \vec{Y}_0 = \begin{bmatrix} u & v & w \end{bmatrix}^t - \begin{bmatrix} -\sqrt{6} & 0 & \sqrt{3} \end{bmatrix}^t, \end{aligned}$$

donde $\vec{Y} = Q^t\vec{X}$ y $\vec{Y}_0 = Q^t\vec{X}_0$, entonces la expresión (5.78) se puede escribir en términos de las nuevas variables $\widetilde{x},\ \widetilde{y},\ \widetilde{z}$ como $\vec{\widetilde{X}}^t\, D\vec{\widetilde{X}}$, o sea,

$$\begin{aligned} 12\widetilde{x}^2 \quad + 6\widetilde{y}^2 + \quad 6\widetilde{z}^2 \quad &= 114 \\ 12\Big(u+\sqrt{6}\Big)^2 + 6v^2 + 6\Big(w-\sqrt{3}\Big)^2 &= 114, \end{aligned} \tag{5.79}$$

la cual corresponde a la ecuación de un elipsoide con centro en $\big(-\sqrt{6}, 0, \sqrt{3}\big)$. Por lo tanto, la ecuación (5.76) es un elipsoide rotado y trasladado.

2. Girar los ejes, dándoles las direcciones principales de la curva y después efectuar una traslación de los ejes girados. Este método se puede emplear cuando A es singular, y para ello se define

$$\vec{Y} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = Q^t\vec{X} = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & 1 & 2 \\ \sqrt{3} & \sqrt{3} & 0 \\ -\sqrt{2} & \sqrt{2} & -\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{5.80}$$

Al sustituir en (5.77) el cambio de variable propuesto en (5.80), se tiene

$$\begin{aligned} 2\vec{K}^t\vec{X} \ &= 2\vec{K}^tQ\vec{Y} = \frac{1}{\sqrt{6}}[-12 \quad 12 \quad 60] \begin{bmatrix} -1 & \sqrt{3} & -\sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \\ 2 & 0 & -\sqrt{2} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \\ &= \begin{bmatrix} \ 24\sqrt{6} & 0 & -12\sqrt{3} \ \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}. \end{aligned}$$

Luego, la ecuación cuadrática dada en (5.76) se puede escribir como

$$12u^2 + 6v^2 + 6w^2 + 24\sqrt{6}u - 12\sqrt{3}w = 24.$$

Al dividir por 6 y completar los cuadrados, se obtiene

$$2\big[u^2 + 2\sqrt{6}u + 6\big] + v^2 + \big[w^2 - 2\sqrt{3}w + 3\big] = 4 + 12 + 3,$$

o bien,

$$2\Big[u+\sqrt{6}\Big]^2 + v^2 + \Big[w-\sqrt{3}\Big]^2 = 19.$$

Nótese que esta superficie es la misma que se obtuvo en (5.79).

5.6.3 Rotación de ejes en $\mathbb{R}^3$

La interpretación geométrica dada a las transformaciones ortogonales en $\mathbb{R}^2$ puede generalizarse para $\mathbb{R}^3$. Esto es, si $\vec{Y} = Q^t\vec{X}$ representa una transformación ortogonal de coordenadas en $\mathbb{R}^3$, entonces las columnas de la matriz de rotación Q están dadas por los cosenos directores de los nuevos ejes de referencia con respecto al viejo sistema de referencia.

Definición 5.16 Rotación de ejes en $\mathbb{R}^3$

Una rotación de ejes en $\mathbb{R}^3$ es una transformación ortogonal propia que permite pasar a una nueva base a partir de un movimiento rígido y continuo de los vectores base del sistema primitivo, conservando el origen fijo y preservando la ortogonalidad.

Teorema 5.22 Rotación de ejes en $\mathbb{R}^3$

Sean $\mathfrak{B} = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ una base del sistema coordenado $\mathbf{X}$ y $\mathfrak{B}' = \{\vec{e}_1', \vec{e}_2', \vec{e}_3'\}$ la base correspondiente al sistema $\mathbf{Y}$. Entonces si:

$$\vec{e}_j' = a_{1j}\vec{e}_1 + a_{2j}\vec{e}_2 + a_{3j}\vec{e}_3, \qquad \text{para} \qquad j = 1, 2, 3,$$

las coordenadas (x_1, x_2, x_3) de un punto cualquiera P en el sistema $\mathbf{X}$ y las coordenadas (y_1, y_2, y_3) del mismo punto en el sistema $\mathbf{Y}$ están relacionadas como sigue:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = A \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix},$$

donde $A = [a_{ij}]$ es la matriz de transición (o matriz de cambio de base) de $\mathfrak{B}$ a $\mathfrak{B}'$.

Demostración.

Queda como ejercicio para el lector. ■

5.6.3.1 Cambio de dirección de ejes en $\mathbb{R}^3$ conservando el mismo origen

Consideremos que el sistema coordenado $\mathbf{X}$ (en $\mathbb{R}^3$) es rectangular y tiene unidades iguales sobre sus tres ejes, lo que significa que $\vec{e}_1$, $\vec{e}_2$ y $\vec{e}_3$, que son los vectores base (unitarios), son perpendiculares entre sí.

Sea $\vec{e}_1'$, $\vec{e}_2'$ y $\vec{e}_3'$ la base ortonormal del sistema $\mathbf{Y}$. Denotemos los productos puntos entre las dos bases por

$$a_{ij} = \langle \vec{e}_i, \vec{e}_j' \rangle.$$

Por ejemplo, $a_{1j} = \langle \vec{e}_1, \vec{e}_j' \rangle$, $a_{2j} = \langle \vec{e}_2, \vec{e}_j' \rangle$ y $a_{3j} = \langle \vec{e}_3, \vec{e}_j' \rangle$ son las tres componentes de $\vec{e}_j'$ con respecto a la base anterior $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ y podemos poner

$$\vec{e}_j' = a_{1j}\vec{e}_1 + a_{2j}\vec{e}_2 + a_{3j}\vec{e}_3, \qquad j = 1, 2, 3.$$

Como $\vec{e}_j'$ es también vector unitario, se tiene que

$$a_{ij} = \cos \sphericalangle(\vec{e}_i, \vec{e}_j') \tag{5.81}$$

y

$$a_{1j}^2 + a_{2j}^2 + a_{3j}^2 = 1, \qquad j = 1, 2, 3.$$

Además, como $\vec{e}_1'$, $\vec{e}_2'$, $\vec{e}_3'$ son ortogonales por pares, es decir $\langle \vec{e}_i', \vec{e}_j' \rangle = 0$, $i \neq j$, se obtiene que

$$a_{1j}a_{1k} + a_{2j}a_{2k} + a_{3j}a_{3k} = 0, \qquad 1 \leq j < k \leq 3.$$

5.6.3.2 Fórmulas de Euler

Euler introdujo en mecánica del sólido rígido unas fórmulas que permiten fijar la posición del segundo sistema coordenado con relación al primero, empleando únicamente tres coordenadas angulares, las cuales se denominan *ángulos de Euler*.

Los movimientos resultantes de variar uno de los ángulos de Euler dejando fijos los otros dos, se llaman "rotaciones de Euler" y tienen nombres particulares: *precesión, rotación intrínseca* y *nutación*.

Estas rotaciones no son ni intrínsecas ni extrínsecas en su totalidad, sino una mezcla de ambas nociones. La precesión es extrínseca, la rotación

intrínseca naturalmente intrínseca, y la nutación es una rotación intermedia alrededor de la intersección de los planos elegidos, la cual se denomina "línea de nodos". Estas rotaciones son commutativas.

Sean X_1, X_2, X_3 los ejes del sistema coordenado $\mathbf{X}$ y representemos por Y_1, Y_2, Y_3 los ejes del sistema de referencia móvil. Se desea definir una transformación ortogonal $\mathbf{Y} = Q^t\mathbf{X}$, que exprese las coordenadas (y_1, y_2, y_3) de un punto arbitrario en términos de sus coordenadas iniciales (x_1, x_2, x_3). Para formular analíticamente esta transformación, se pasará del sistema de referencia de ejes ortogonales $\mathbf{X}$ a otro sistema de referencia con origen común $\mathbf{Y}$ mediante tres cambios sucesivos, conservando la ortogonalidad. En cada cambio, se asumirá que se tiene en común con el sistema precedente un eje y el plano opuesto. De esta manera, no se exigirán más que las fórmulas correspondientes al cambio de ejes situados en el mismo plano.

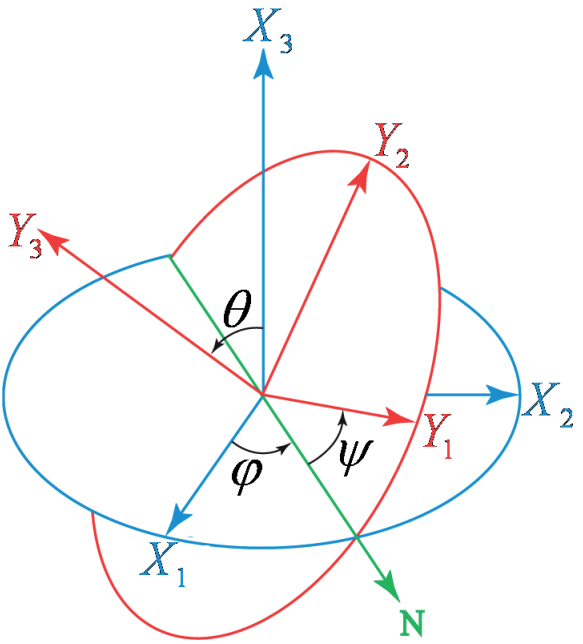

Figura 5.4: Rotación en $\mathbb{R}^3$.

1^0. Se obtienen unos nuevos ejes $W_1, W_2, W_3 \equiv X_3$ rotando el plano que contienen los vectores $\vec{e}_1$ y $\vec{e}_2$ del sistema $\mathbf{X}$ un ángulo φ alrededor del eje X_3, en sentido contrario al de las manecillas del reloj. Por el

Teorema 5.22, se tiene que las coordenadas de un punto en ambos sistemas están relacionadas por

$$\begin{aligned} x_1 &= w_1 \cos\varphi - w_2 \operatorname{sen}\varphi, \\ x_2 &= w_1 \operatorname{sen}\varphi + w_2 \cos\varphi, \\ x_3 &= w_3. \end{aligned} \tag{5.82}$$

Al expresar dicha rotación en forma matricial, se obtiene

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} \cos\varphi & \operatorname{sen}\varphi & 0 \\ -\operatorname{sen}\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{R}_\varphi \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

2^0. Se generan otros nuevos ejes $Z_1 \equiv W_1$, Z_2, Z_3 rotando el plano determinado por los ejes W_2 y W_3 un ángulo θ alrededor del eje W_1, lo que hará tomar al eje X_3 la posición Y_3 y al W_2 la Z_2. Luego, por el Teorema 5.22, se deduce que las coordenadas de transformación de un punto, serán

$$\begin{aligned} w_1 &= z_1, \\ w_2 &= z_2 \cos\theta - z_3 \operatorname{sen}\theta, \\ w_3 &= z_2 \operatorname{sen}\theta + z_3 \cos\theta. \end{aligned} \tag{5.83}$$

Si se expresa (5.83) en forma matricial, se tiene que

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \operatorname{sen}\theta \\ 0 & -\operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \mathbf{R}_\theta \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}.$$

3^0. Finalmente, se rota alrededor del eje $Z_3 \equiv Y_3$ el plano que contiene a los dos ejes Z_2 y Z_3 hasta que formen un ángulo ψ, así quedará W_1 en Y_1 y Z_2 en Y_2 y por el Teorema 5.22, las coordenadas de un punto en ambos sistemas están relacionadas por

$$\begin{aligned} z_1 &= y_1 \cos\psi - y_2 \operatorname{sen}\psi, \\ z_2 &= y_1 \operatorname{sen}\psi + y_2 \cos\psi, \\ z_3 &= y_3. \end{aligned} \tag{5.84}$$

Al expresar (5.84) en forma matricial, se obtiene

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \cos\psi & \operatorname{sen}\psi & 0 \\ -\operatorname{sen}\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \mathbf{R}_\psi \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.$$

La eliminación de los sistemas coordenados **W** y **Z** en las ecuaciones (5.82), (5.83) y (5.84) dará la transformación del producto $\mathbf{Y} = \mathbf{RX}$, mediante la matriz $\mathbf{R} = \mathbf{R}_\psi \mathbf{R}_\theta \mathbf{R}_\varphi$ de rotación en $\mathbb{R}^3$ que representa la transformación de coordenadas desde el sistema fijo al sistema móvil,

$$\mathbf{R} = \begin{bmatrix} \cos\psi & \operatorname{sen}\psi & 0 \\ -\operatorname{sen}\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \operatorname{sen}\theta \\ 0 & -\operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\varphi & \operatorname{sen}\varphi & 0 \\ -\operatorname{sen}\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} \cos\psi\cos\varphi - \operatorname{sen}\psi\operatorname{sen}\varphi\cos\theta & \operatorname{sen}\psi\cos\varphi\cos\theta + \operatorname{sen}\varphi\cos\psi & \operatorname{sen}\psi\operatorname{sen}\theta \\ -\operatorname{sen}\psi\cos\varphi - \operatorname{sen}\varphi\cos\psi\cos\theta & \cos\psi\cos\varphi\cos\theta - \operatorname{sen}\psi\operatorname{sen}\varphi & \cos\psi\operatorname{sen}\theta \\ \operatorname{sen}\varphi\operatorname{sen}\theta & -\cos\varphi\operatorname{sen}\theta & \cos\theta \end{bmatrix};$$

los ángulos $\varphi,\ \theta,\ \psi$ deben escogerse en el intervalo $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ y se conocen como *ángulos eulerianos*. Nótese que el determinante de esta matriz es

$$\det(\mathbf{R}) = \det(\mathbf{R}_\psi \mathbf{R}_\theta \mathbf{R}_\varphi) \ = \ \det(\mathbf{R}_\psi)\det(\mathbf{R}_\theta)\det(\mathbf{R}_\varphi) = 1.$$

Luego, esta matriz es ortogonal.

La comparación de los elementos de la matriz $\mathbf{R}_\psi \mathbf{R}_\theta \mathbf{R}_\varphi$ con las expresiones dadas en (5.81) permiten obtener esos nueve cosenos en función de las constantes $\varphi,\ \theta$ y ψ.

Observaciones

Para determinar los ángulos eulerianos $\varphi,\ \theta$ y ψ, se comparan los elementos de $Q^t = [c_{ij}]$ con los de $\mathbf{R}_\psi \mathbf{R}_\theta \mathbf{R}_\varphi$ (donde las columnas de Q son los vectores propios normalizados de la matriz simétrica asociada a la forma cuadrática) y se tiene en cuenta que:

1. Si $c_{33} \neq 1$, entonces los ángulos eulerianos se pueden determinar de la siguiente manera:

$$\cos\theta \ = c_{33}, \qquad \tan\psi \ = \frac{c_{13}}{c_{23}} \qquad \text{y} \qquad \tan\varphi \ = \ -\frac{c_{31}}{c_{32}}.$$

2. Cuando $c_{33} = 1$, entonces $\theta = 0$ y la matriz de rotación $\mathbf{R}_\psi \mathbf{R}_\theta \mathbf{R}_\varphi$ tiene la forma

$$\mathbf{R}_\psi \mathbf{R}_\theta \mathbf{R}_\varphi = \begin{bmatrix} \cos\big(\psi+\varphi\big) & \operatorname{sen}\big(\psi+\varphi\big) & 0 \\ -\operatorname{sen}\big(\psi+\varphi\big) & \cos\big(\psi+\varphi\big) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

en este caso,

$$\tan\big(\psi+\varphi\big) = -\frac{c_{21}}{c_{11}}.$$

Aquí, los ángulos ψ y φ se escogen arbitrariamente.

3. Si la suma de los cuadrados de los elementos de la diagonal principal de la matriz de rotación $\mathbf{R}_\psi\mathbf{R}_\theta\mathbf{R}_\varphi$ es igual a 1, entonces, el ángulo θ satisface que

$$\cos\theta = \frac{2\tan\varphi\tan\psi \pm \sqrt{\left(\sec^2\varphi + \sec^2\psi\right)\left(\sec^2\varphi\tan^2\psi + \sec^2\psi\tan^2\varphi\right)}}{1 + \tan^2\varphi\tan^2\psi + \sec^2\varphi\sec^2\psi}.$$

En el último caso, como la tangente de $\frac{\pi}{2}$ no está definida, cuando uno de los ángulos φ ó ψ sea igual a $\pm\frac{\pi}{2}$, se tiene que

$$\cos\theta = \begin{cases} \dfrac{1}{\sqrt{\tan^2\psi + \sec^2\psi}} & \text{si } \varphi \to \left(\pm\frac{\pi}{2}\right) \\ \dfrac{1}{\sqrt{\tan^2\varphi + \sec^2\varphi}} & \text{si } \psi \to \left(\pm\frac{\pi}{2}\right), \end{cases}$$

y si uno de los ángulos φ ó ψ es igual a cero, entonces

$$\cos\theta = \begin{cases} \dfrac{\left|\tan\psi\right|}{\sqrt{1 + \sec^2\psi}} & \text{si } \varphi = 0 \\ \dfrac{\left|\tan\varphi\right|}{\sqrt{1 + \sec^2\varphi}} & \text{si } \psi = 0. \end{cases}$$

Ejemplo 5.23 *Determine los ángulos eulerianos de la ecuación cuadrática dada en el Ejemplo 5.21.*

Solución.

La matriz ortogonal asociada a la forma cuadrática era

$$Q = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}.$$

Al comparar los elementos de Q^t con las de $\mathbf{R}_\psi\mathbf{R}_\theta\mathbf{R}_\varphi$, se tiene que

$$\cos\theta = \frac{1}{\sqrt{6}}, \qquad \tan\psi = \sqrt{\frac{3}{2}} \qquad \text{y} \qquad \tan\varphi = -\frac{1}{2}. \tag{5.85}$$

Se debe escoger una solución de (5.85) para la cual los ángulos eulerianos se encuentren dentro del intervalo $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. En este caso, los ángulos θ, ψ y φ valen, aproximadamente, $65°54'18.57''$, $50°46'6.53''$ y $-26°33'54.18''$.

Ejemplo 5.24 *Determine los ángulos eulerianos de la ecuación cuadrática dada en el Ejemplo 5.22.*

Solución.

La matriz ortogonal asociada a la forma cuadrática era

$$Q = \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \end{bmatrix}.$$

Si se comparan los elementos de Q^t con los de $\mathbf{R}_\psi \mathbf{R}_\theta \mathbf{R}_\varphi$, se tiene que

$$\cos\theta = -\frac{1}{\sqrt{3}}, \qquad \tan\psi = \textit{indefinida} \qquad \text{y} \qquad \tan\varphi = 1. \tag{5.86}$$

Se debe escoger una solución de (5.86) para la cual los ángulos eulerianos se encuentren dentro del intervalo $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, en este caso, los ángulos θ, ψ y φ valen, aproximadamente, $125°15'51.8''$, $90°$ y $45°$.

5.6.4 Clasificación de las superficies cuádricas

Para la superficie cuádrica

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0, \tag{5.87}$$

se establecen las siguientes cantidades

$$A = \begin{bmatrix} a & \frac{d}{2} & \frac{e}{2} \\ \frac{d}{2} & b & \frac{f}{2} \\ \frac{e}{2} & \frac{f}{2} & c \end{bmatrix}, \qquad \begin{array}{c} \omega = \operatorname{tr}(A), \\ \mu = \sum_{i=1}^{3} M_{ii}(A), \\ \nu = \det A, \end{array} \qquad \delta = \det \begin{bmatrix} A & \vdots & \vec{K} \\ \cdots & \cdot & \cdots \\ \vec{K}^t & \vdots & j \end{bmatrix},$$

donde $\vec{K}^t = \begin{bmatrix} \frac{g}{2} & \frac{h}{2} & \frac{i}{2} \end{bmatrix}$ y ω, μ, ν son los coeficientes de $P_A(\lambda)$. Luego, si $\lambda_1, \lambda_2, \lambda_3$ son los valores propios de A, entonces la ecuación cuádrica (5.87) representa los siguientes lugares geométricos (L.G.):

$$
\text{L.G.}\begin{cases}
\nu \neq 0 \begin{cases}
\dfrac{\delta}{\nu} \neq 0 \ \left[\dfrac{\delta}{\nu} < 0\right] \begin{cases}
\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0 & \textbf{Un elipsoide} \\
\lambda_1 > 0, \lambda_2 > 0, \lambda_3 < 0 & \textbf{Un hiperboloide de una hoja} \\
\lambda_1 > 0, \lambda_2 < 0, \lambda_3 < 0 & \textbf{Un hiperboloide de dos hojas} \\
\lambda_1 < 0, \lambda_2 < 0, \lambda_3 < 0 & \textbf{Conjunto vacío}
\end{cases} \\
\dfrac{\delta}{\nu} = 0 \begin{cases}
\lambda_1, \lambda_2, \lambda_3 \text{ tienen el mismo signo.} & \textbf{Un punto} \\
\lambda_1, \lambda_2, \lambda_3 \text{ tienen signos distintos.} & \textbf{Un cono}
\end{cases}
\end{cases} \\
\nu = 0 \ [\lambda_3 = 0] \begin{cases}
\widetilde{I} = 0 \begin{cases}
\dfrac{\delta}{\nu'} \neq 0 \begin{cases}
\lambda_1 > 0, \lambda_2 > 0 & \textbf{Un cilindro elíptico} \\
\lambda_1 > 0, \lambda_2 < 0 & \textbf{Un cilindro hiperbólico} \\
\lambda_1 < 0, \lambda_2 < 0 & \textbf{Conjunto vacío}
\end{cases} \\
\dfrac{\delta}{\nu'} = 0 \begin{cases}
\lambda_1\lambda_2 > 0 & \textbf{Una recta} \\
\lambda_1\lambda_2 < 0 & \textbf{Dos planos que se cortan}
\end{cases}
\end{cases} \\
\widetilde{I} \neq 0 \begin{cases}
\lambda_1\lambda_2 > 0 & \textbf{Un paraboloide elíptico} \\
\lambda_1\lambda_2 < 0 & \textbf{Un paraboloide hiperbólico}
\end{cases}
\end{cases} \\
\nu = 0 \ [\lambda_1 \neq 0] \begin{cases}
\widetilde{I} = 0 \text{ y } \widetilde{H} = 0 \begin{cases}
\dfrac{\delta}{\nu^*}\lambda_1 > 0 & \textbf{Conjunto vacío} \\
\dfrac{\delta}{\nu^*}\lambda_1 < 0 & \textbf{Dos planos paralelos}
\end{cases} \\
\dfrac{\delta}{\nu^*} \neq 0 \text{ y/o } \widetilde{H} \neq 0 \quad \textbf{Un cilindro parabólico,}
\end{cases}
\end{cases}
$$

donde

$$
\frac{\delta}{\nu} = j - \sum_{i=1}^{3} \frac{1}{\lambda_i}\big(\vec{K}^t\vec{w}_i\big)^2, \qquad \frac{\delta}{\nu'} = j - \sum_{i=1}^{2} \frac{1}{\lambda_i}\big(\vec{K}^t\vec{w}_i\big)^2,
$$

$$
\widetilde{I} = \vec{K}^t\vec{w}_3, \qquad \widetilde{H} = \vec{K}^t\vec{w}_2,
$$

aquí $\vec{w}_i$ es el vector propio de A asociado al valor propio λ_i. Cuando $\delta \neq 0$, estas superficies pueden ser clasificadas teniendo en cuenta la inercia de la matriz simétrica A como sigue:

Identificación de las superficies cuádricas

Inercia	**Nombre de la superficie**
$In(A) = (3, 0, 0)$	Elipsoide
$In(A) = (2, 1, 0)$	Hiperboloide de una hoja
$In(A) = (1, 2, 0)$	Hiperboloide de dos hojas
$In(A) = (2, 0, 1)$	Paraboloide elíptico
$In(A) = (1, 1, 1)$	Paraboloide hiperbólico
$In(A) = (1, 0, 2)$	Cilindro parabólico

Ejercicios 5.5

1. Determine la sección cónica y el ángulo de rotación para:

   a. $4x^2 + 4xy + y^2 = 9$.

   b. $36x^2 + 9y^2 + 4z^2 - 36 = 0$.

   c. $x^2 + 8xy + 16y^2 - 4x + 16y = -7$.

   d. $x^2 + 2xy + y^2 + 2x + 2y = -1$.

   e. $4x^2 - 20xy + 25y^2 + 4x - 10y = -1$.

   f. $4x^2 - 4xy + y^2 - 6x + 3y = 4$.

   g. $9x^2 - 24xy + 16y^2 - 20x + 110y = 50$.

   h. $6x^2 + 3y^2 - 2z^2 + 12x - 18y - 8z = -7$.

   i. $2x^2 + 2y^2 + 5z^2 - 4xy + 2xz - 2yz + 10x - 26y - 2z = 0$.

2. Sea A la representación matricial de la ecuación cuadrática (5.58) con $f \neq 0$. Sean λ_1 y λ_2 los valores propios de A. Demuestre que la curva que describe (5.58) es:

   a. Una hipérbola si $\lambda_1\lambda_2 < 0$.

   b. Un par de rectas o una parábola si $\lambda_1\lambda_2 = 0$.

   c. Un circulo, elipse o sección cónica degenerada si $\lambda_1\lambda_2 > 0$.

3. Sean:

$$A = \begin{bmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} x & \frac{y}{2} \\ \frac{y}{2} & z \end{bmatrix},$$

   demuestre que si $\frac{b}{a-c} = \frac{y}{x-z}$, entonces $AB = BA$.

Capítulo 6

Formas hermíticas

En el capítulo anterior se desarrolló la teoría para formas cuadráticas con matriz asociada simétrica real. En este capítulo se consideran formas cuadráticas pero con *matriz asociada compleja*. Se estudia el caso *complejo* independientemente del caso *real*, ya que si se asume $\vec{X} \in \mathbb{C}^2$ y se obtiene la expresión $\|\vec{X}\|^2 = \vec{X}^t\vec{X}$, de manera análoga al producto escalar estándar de $\mathbb{R}^2$, se llega a resultados ilógicos. Por ejemplo, para el vector *no nulo* $\vec{X}^t = \big(a, bi\big)$, se tiene que

$$\vec{X}^t\vec{X} = a^2 + b^2i^2 = a^2 - b^2;$$

este producto puede ser cero si $a = b$ ó $a = -b$, hecho que contradice la propiedad (v) del producto escalar estándar en $\mathbb{R}^n$ (ver Capítulo 1). Este hecho induce a la redefinición de formas cuadráticas para el caso complejo.

6.1 Forma hermítica

Definición 6.1 Forma sesquilineal

Sea $\mathbb{V}$ un espacio vectorial de dimensión finita sobre el cuerpo complejo $\mathbb{C}$. Una forma sesquilineal es una función $g : \mathbb{V} \times \mathbb{V} \to \mathbb{C}$ tal que

$$(i)\ g\big(\alpha\vec{u}_1 + \beta\vec{u}_2, \vec{v}\big) = \overline{\alpha}g\big(\vec{u}_1, \vec{v}\big) + \overline{\beta}g\big(\vec{u}_2, \vec{v}\big)$$

$$(ii)\ g\big(\vec{u}, \alpha\vec{v}_1 + \beta\vec{v}_2\big) = \alpha g\big(\vec{u}, \vec{v}_1\big) + \beta g\big(\vec{u}, \vec{v}_2\big),$$

donde $\alpha, \beta \in \mathbb{C}$ *y* $\vec{u}_i, \vec{v}_j \in \mathbb{V}$.

Como en la Definición 5.1, la condición (ii) se interpreta como que g es lineal en la segunda variable. Por otra parte, expresamos la condición (i) diciendo que g es *lineal conjugada* en la primera variable. En el resto de esta sección se omitirá el adjetivo "sesquilineal", salvo que sea importante tenerlo en cuenta.

Definición 6.2 *Sea* $g : \mathbb{V} \times \mathbb{V} \to \mathbb{C}$ *una forma sobre* $\mathbb{V}$*, entonces* g *es hermítica, si para todo* $\vec{v},\ \vec{w} \in \mathbb{V}$*, se cumple que:*

$$g\big(\vec{v}, \vec{w}\big) = \overline{g(\vec{w}, \vec{v})}. \tag{6.1}$$

Ejemplo 6.1 *Sea* $g : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{R}$ *definida por*

$$g\big(\vec{X}, \vec{Y}\big) = \vec{X}^H A \vec{Y},$$

donde $\vec{X}$ *y* $\vec{Y} \in \mathbb{C}^n$ *y* A *es una matriz hermitiana. Verifique si la aplicación* g *define una forma hermítica sobre* $\mathbb{C}^n$.

Solución.

(i) Para todo $\vec{X}_1,\ \vec{X}_2$ y $\vec{Y} \in \mathbb{C}^n$, se tiene que

$$\begin{aligned} g\big(\vec{X}_1 + \vec{X}_2, \vec{Y}\big) &= \big(\vec{X}_1 + \vec{X}_2\big)^H A\vec{Y} = \big(\vec{X}_1^H + \vec{X}_2^H\big)A\vec{Y} \\ &= \vec{X}_1^H A\vec{Y} + \vec{X}_2^H A\vec{Y} = g\big(\vec{X}_1, \vec{Y}\big) + g\big(\vec{X}_2, \vec{Y}\big). \end{aligned}$$

(ii) Para todo $\beta \in \mathbb{C}$, $\vec{X}$ y $\vec{Y} \in \mathbb{C}^n$, se tiene que

$$\begin{aligned} g\big(\vec{X}, \beta\vec{Y}\big) &= \vec{X}^H A\big(\beta\vec{Y}\big) = \beta\vec{X}^H A\vec{Y} \\ &= \beta\vec{X}^H A\vec{Y} = \beta g\big(\vec{X}, \vec{Y}\big). \end{aligned}$$

Así, la aplicación g es lineal en la segunda variable. Además,

$$\overline{g\big(\vec{X}, \vec{Y}\big)} = \overline{\big(\vec{X}^H A\vec{Y}\big)} = \overline{\big(\vec{X}^H A\vec{Y}\big)^t} = \vec{Y}^H A^H \vec{X} = \vec{Y}^H A\vec{X} = g\big(\vec{Y}, \vec{X}\big).$$

Por lo tanto, g es una forma hermítica sobre $\mathbb{C}^n$.

Teorema 6.1 *Sean $\mathbb{V}$ un espacio vectorial complejo y g una forma sesquilineal sobre $\mathbb{V}$ tal que $g(\vec{u},\vec{u})$ sea real para todo $\vec{u}\in\mathbb{V}$. Entonces, g es hermítica.*

Demostración.

Sean $\vec{u},\vec{v}\in\mathbb{V}$ y g una forma sesquilineal sobre $\mathbb{V}$ tal que $g(\vec{u},\vec{u})$ sea real para todo $\vec{u}\in\mathbb{V}$. Se debe probar que $g(\vec{u},\vec{v})=\overline{g(\vec{v},\vec{u})}$. En efecto,

$$g(\vec{u}+\vec{v},\vec{u}+\vec{v})=g(\vec{u},\vec{u})+g(\vec{u},\vec{v})+g(\vec{v},\vec{u})+g(\vec{v},\vec{v}).$$

Como por hipótesis $g(\vec{u}+\vec{v},\vec{u}+\vec{v})$, $g(\vec{u},\vec{u})$ y $g(\vec{v},\vec{v})$ son reales, el número $g(\vec{u},\vec{v})+g(\vec{v},\vec{u})$ es real. De manera análoga, se tiene

$$g(\vec{u}+i\vec{v},\vec{u}+i\vec{v})=g(\vec{u},\vec{u})+ig(\vec{u},\vec{v})-ig(\vec{v},\vec{u})+g(\vec{v},\vec{v}).$$

Por el mismo razonamiento anterior, vemos que $ig(\vec{u},\vec{v})-ig(\vec{v},\vec{u})$ es real. Al concluir que estos dos números son reales, se pueden igualar sus complejos conjugados y se obtiene

$$g(\vec{u},\vec{v})+g(\vec{v},\vec{u}) = \overline{g(\vec{u},\vec{v})}+\overline{g(\vec{v},\vec{u})} \tag{6.2}$$

$$ig(\vec{u},\vec{v})-ig(\vec{v},\vec{u}) = -i\overline{g(\vec{u},\vec{v})}+i\overline{g(\vec{v},\vec{u})}. \tag{6.3}$$

Al multiplicar (6.3) por $(-i)$ y sumarla a la expresión (6.2), se llega a

$$2g(\vec{u},\vec{v})=2\overline{g(\vec{v},\vec{u})}.$$

Por lo tanto, se satisface (6.1). ■

Teorema 6.2 *Sea $\mathbb{V}$ un espacio vectorial de dimensión finita sobre $\mathbb{C}$. Sea g una forma hermitiana sobre $\mathbb{V}$. Entonces, existe una matriz única hermitiana A, tal que para todo $\vec{u},\vec{v}\in\mathbb{V}$,*

$$g_A(\vec{u},\vec{v})=\vec{u}^{H}A\vec{v}. \tag{6.4}$$

Demostración.

La prueba es completamente análoga a la del Teorema 5.1 y se deja como ejercicio para el lector. ■

Teorema 6.3 Identidad de polarización

Sea g una forma hermitiana sobre un espacio vectorial complejo $\mathbb{V}$, entonces para todo $\vec{u}, \vec{v} \in \mathbb{V}$, se cumple que

$$g_A\big(\vec{u}+\vec{v}, \vec{u}+\vec{v}\big) - g_A\big(\vec{u}-\vec{v}, \vec{u}-\vec{v}\big) = 2\big[g_A(\vec{u}, \vec{v}) + g_A(\vec{v}, \vec{u})\big]. \quad (6.5)$$

Demostración.

La verificación de esta identidad se hace en forma trivial, solo desarrollando el miembro izquierdo que aparece en (6.5). ■

Teorema 6.4 *Sea $\mathbb{V}$ como antes. Si g es una forma hermitiana tal que $g_A(\vec{v}, \vec{v}) = 0$ para todo $\vec{v} \in \mathbb{V}$, entonces $A = \mathrm{O}$.*

Demostración.

Por el Teorema 6.3, para todo $\vec{u}, \vec{v} \in \mathbb{V}$, se tiene que

$$g_A(\vec{u}+\vec{v}, \vec{u}+\vec{v}) - g_A(\vec{u}-\vec{v}, \vec{u}-\vec{v}) = 2\big[g_A(\vec{u}, \vec{v}) + g_A(\vec{v}, \vec{u})\big].$$

Luego, si g es tal que $g_A(\vec{v}, \vec{v}) = 0$ para todo $\vec{v} \in \mathbb{V}$, el miembro izquierdo de la identidad de polarización es igual a 0, de donde se obtiene que

$$g_A(\vec{u}, \vec{v}) + g_A(\vec{v}, \vec{u}) = 0, \quad (6.6)$$

para todo $\vec{u}, \vec{v} \in \mathbb{V}$. Si se reemplaza $\vec{u}$ por $i\vec{u}$, entonces se tiene que

$$g_A\big(i\vec{u}, \vec{v}\big) + g_A\big(\vec{v}, i\vec{u}\big) = -ig_A\big(\vec{u}, \vec{v}\big) + ig_A\big(\vec{v}, \vec{u}\big) = 0.$$

Así,

$$-g_A(\vec{u}, \vec{v}) + g_A(\vec{v}, \vec{u}) = 0. \quad (6.7)$$

Si se restan las relaciones (6.6) y (6.7), se obtiene

$$2g_A(\vec{u}, \vec{v}) = 0.$$

Por lo tanto, $g_A(\vec{u}, \vec{v}) = 0$. Luego, $A = \mathrm{O}$. Como se quería demostrar. ■

Teorema 6.5 *Una matriz compleja A de tamaño $n \times n$ representa una forma hermítica si y solo si es una matriz hermitiana.*

Demostración.

Supóngase que A es hermitiana. Como para todo $\vec{X},\ \vec{Y} \in \mathbb{C}^n$, la matriz $\vec{X}^H A\vec{Y}$ es una matriz de 1×1, es decir, un elemento de $\mathbb{R}$, entonces es igual a su propia transpuesta conjugada. Por lo tanto,

$$\overline{\vec{X}^H A\vec{Y}} = \overline{\left(\vec{X}^H A\vec{Y}\right)^t} = \overline{\vec{Y}^t A^t \overline{\vec{X}}} = \vec{Y}^H A^H \vec{X} = \vec{Y}^H A\vec{X},$$

así que A representa una forma hermitiana.

Recíprocamente, supóngase que A representa una forma hermítica; es decir,

$$g_A\left(\vec{X},\vec{Y}\right) = \overline{g_A\left(\vec{Y},\vec{X}\right)}, \tag{6.8}$$

para todo $\vec{X},\ \vec{Y} \in \mathbb{C}^n$. Como

$$\overline{g_A\left(\vec{Y},\vec{X}\right)} = \overline{\vec{Y}^H A\vec{X}} = \overline{\left(\vec{Y}^H A\vec{X}\right)^t} = \vec{X}^H A^H \vec{Y}, \tag{6.9}$$

al comparar (6.8) y (6.9), se tiene que

$$g_A\left(\vec{X},\vec{Y}\right) = \vec{X}^t A\vec{Y} = \vec{X}^t A^H \vec{Y}. \tag{6.10}$$

Como (6.10) se cumple para todo $\vec{X},\ \vec{Y} \in \mathbb{C}^n$, se concluye que $A = A^H$, es decir, A es hermitiana. ■

6.2 Forma cuadrática compleja

En esta sección se estudian las formas cuadráticas $F(\vec{X}) = \vec{X}^H A\vec{X}$, en donde A es una matriz compleja de tamaño $n \times n$ y la variable $\vec{X}$ se escoge en $\mathbb{C}^n$. Como en la práctica, generalmente uno solo se preocupa de las formas cuadráticas $F(\vec{X}) = \vec{X}^H A\vec{X}$ que toman únicamente valores reales, en este apartado se asumirá que la matriz A asociada a la forma es hermitiana. Cabe notar que en los casos en que $F(\vec{X})$ es compleja, por lo general solo se puede estudiar la parte real de $F(\vec{X})$.

Definición 6.3 Forma cuadrática compleja

Sea $\mathbb{V}$ un espacio vectorial de dimensión finita sobre los números complejos. Sea $g : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ una forma hermítica sobre $\mathbb{V}$. Entonces, una

forma cuadrática hermítica o forma cuadrática compleja determinada por g *es una función* $F : \mathbb{V} \to \mathbb{R}$*, tal que*

$$F(\vec{v}) = g_A(\vec{v}, \vec{v}) = \vec{v}^H A \vec{v}. \tag{6.11}$$

La matriz A *se llama la representación matricial de la forma cuadrática compleja.*

Ejemplo 6.2 Producto hermítico canónico

Sea $\mathbb{V} = \mathbb{C}^n$ *y considere la forma cuadrática compleja determinada por el producto escalar sobre* $\mathbb{C}^n$*:*

$$F(\vec{v}) = |z_1|^2 + |z_2|^2 + \ldots + |z_n|^2,$$

donde $|z_i|^2 = \overline{z_i} z_i$*. Exprese esta forma cuadrática compleja como* $\vec{v}^H A \vec{v}$*.*

Solución.

Vamos a determinar la matriz compleja $A = [a_{ij}]$ de la forma hermítica g, de tal forma que

$$F(\vec{v}) = \vec{v}^H A \vec{v} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \overline{z_i} z_j. \tag{6.12}$$

Es decir, se quieren encontrar los valores de a_{ij}, de manera que

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \overline{z_i} z_j = \sum_{i=1}^{n} |z_i|^2.$$

Como la matriz A es hermitiana, $\overline{a_{ij}} = a_{ji}$. Por lo tanto, la forma cuadrática compleja dada en (6.12) se puede expresar como

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \overline{z_i} z_j = \sum_{i=1}^{n} a_{ii} |z_i|^2 + \sum_{i=1}^{n-1} \sum_{j>i}^{n} a_{ij} \overline{z_i} z_j + \sum_{i=1}^{n-1} \sum_{j>i}^{n} \overline{a_{ij}}\ z_i \overline{z_j}, \tag{6.13}$$

y si se comparan términos, se establecen las siguientes relaciones:

$$\sum_{i=1}^{n} a_{ii} |z_i|^2 \ = \ \sum_{i=1}^{n} |z_i|^2 \quad \text{y} \quad \sum_{i=1}^{n-1} \sum_{j>i}^{n} a_{ij} \overline{z_i} z_j + \sum_{i=1}^{n-1} \sum_{j>i}^{n} \overline{a_{ij}}\ z_i \overline{z_j} \ = 0.$$

Pero como en la función $F(\vec{v})$ no aparecen términos de la forma $\overline{z_i}z_j$, entonces

$$a_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Luego, $A = I_n$ y, por lo tanto, $F(\vec{v})$ se puede expresar como $\vec{v}^H I_n \vec{v}$.

Ejemplo 6.3 *Sea* $\mathbb{V} = \mathbb{C}^3$ *y* $F(\vec{X}) = |x_1|^2 - i\overline{x_1}x_2 + i\overline{x_2}x_1 - \overline{x_1}x_3 - \overline{x_3}x_1 - 2i\overline{x_2}x_3 + 2i\overline{x_3}x_2$. *Exprese esta forma hermítica como* $\vec{X}^H A\vec{X}$.

Solución.

Si se utiliza el resultado obtenido en (6.13), para $n = 3$, se tiene que

$$\vec{X}^H A\vec{X} = \sum_{i=1}^{3} a_{ii}|x_i|^2 + \sum_{i=1}^{2}\sum_{j>i}^{3} a_{ij}\overline{x_i}x_j + \sum_{i=1}^{2}\sum_{j>i}^{3} \overline{a_{ij}}x_i\overline{x_j}.$$

Al resolver esta suma y comparar los a_{ij} con los coeficientes de la función $F(\vec{X})$, se obtiene la matriz

$$A = \begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix},$$

la cual permite expresar a $F(\vec{X})$ de la forma $\vec{X}^H A\vec{X}$.

6.3 Diagonalización de una forma hermítica

Teorema 6.6 *Sea* $F(\vec{X})$ *una forma cuadrática compleja asociada a una matriz hermítica* A. *Sea* L *una matriz compleja triangular inferior tal que* A *se pueda factorizar como* LDL^H. *El cambio de coordenadas*

$$\vec{Z} = L^H\vec{X} \tag{6.14}$$

transforma a $\vec{X}^H A\vec{X}$ *en* $\vec{Z}^H D\vec{Z}$.

Demostración.

La matriz A asociada a la forma se puede factorizar como

$$A = LDU.$$

Como A es hermítica, por el Teorema 3.5, $U = L^H$. Por lo tanto,

$$\begin{aligned} \vec{X}^H A\vec{X} &= \vec{X}^H\big(LDL^H\big)\vec{X} && \text{puesto que } A = LDL^H \\ &= \big(\vec{X}^H L\big)D\big(L^H\vec{X}\big) \\ &= \big(L^H\vec{X}\big)^H D\big(L^H\vec{X}\big) = \vec{Z}^H D\vec{Z} && \text{puesto que } \vec{Z} = L^H\vec{X}. \end{aligned}$$

Así, queda probado el teorema. ■

A continuación, se presenta una versión de este método de diagonalización:

Procedimiento para diagonalizar una forma hermítica

i) Halle la matriz de coeficientes hermítica A asociada a $F(\vec{X})$.

ii) Obtenga la descomposición LDL^H de A sin efectuar intercambios de filas que destruyan el hecho de que $\overline{a_{ij}} = a_{ji}$ y con elementos en $D = diag\{d_{11}, d_{22}, \ldots, d_{nn}\}$ tales que $d_{ii} \in \mathbb{R}$ no necesariamente distintos de cero. Además, $\det(L) = 1$.

iii) Transforme a $F(\vec{X})$ en $d_{11}|z_1|^2 + d_{22}|z_2|^2 + \ldots + d_{nn}|z_n|^2$, bajo el cambio de coordenadas $\vec{Z} = L^H\vec{X}$.

Ejemplo 6.4 *Considere la ecuación cuadrática compleja:*

$$|x_1|^2 - i\overline{x_1}x_2 + i\overline{x_2}x_1 - \overline{x_1}x_3 - \overline{x_3}x_1 - 2i\overline{x_2}x_3 + 2i\overline{x_3}x_2 = 9. \qquad (6.15)$$

Encuentre una diagonalización para esta forma hermítica, usando el método descrito anteriormente.

Solución.

En el Ejemplo 6.3, se obtuvo que la forma cuadrática hermítica asociada

$$|x_1|^2 - i\overline{x_1}x_2 + i\overline{x_2}x_1 - \overline{x_1}x_3 - \overline{x_3}x_1 - 2i\overline{x_2}x_3 + 2i\overline{x_3}x_2$$

se puede expresar matricialmente como

$$\vec{X}^H A\vec{X} = [\overline{x_1} \;\; \overline{x_2} \;\; \overline{x_3}] \begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{6.16}$$

La factorización LDL^H de la matriz asociada a la forma hermítica es

$$\begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ i & 1 & 0 \\ -1 & -i & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -i & -1 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{bmatrix},$$

de modo que

$$\begin{aligned} F(\vec{X}) &= \vec{X}^H \begin{bmatrix} 1 & -i & -1 \\ i & 0 & -2i \\ -1 & 2i & 0 \end{bmatrix} \vec{X} \\ &= \vec{X}^H \begin{bmatrix} 1 & 0 & 0 \\ i & 1 & 0 \\ -1 & -i & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -i & -1 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{bmatrix} \vec{X} \end{aligned}$$

si se hace

$$\vec{Z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = L^H \vec{X} = \begin{bmatrix} 1 & -i & -1 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{bmatrix} \vec{X}.$$

Nótese que $\det(L) = 1$. Por lo tanto, el cambio de variables

$$z_1 = x_1 - ix_2 - x_3, \qquad z_2 = x_2 + ix_3 \qquad \text{y} \qquad z_3 = x_3$$

permite expresar a $F(\vec{X})$ de la siguiente manera:

$$F(\vec{Z}) = |z_1|^2 - |z_2|^2 = 9,$$

y al reescribir (6.15), se tiene que

$$|z_1|^2 - |z_2|^2 = 9,$$

la cual corresponde a la ecuación de una "hipérbola" sobre los ejes z_1 y z_2.

Teorema 6.7 Teorema de los ejes principales

Sea $F(\vec{X})$ una forma hermítica asociada a una matriz hermitiana A con valores propios (no necesariamente distintos) $\lambda_1, \lambda_2, \ldots, \lambda_n$. Sea U una matriz unitaria que diagonaliza a A. El cambio de coordenadas:

$$\vec{X} = U\vec{Z} \tag{6.17}$$

transforma a $\vec{X}^H A\vec{X}$ en $\vec{Z}^H D\vec{Z}$, donde la matriz diagonal

$$D = U^H AU = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\}.$$

Demostración.

La demostración consiste en un cálculo directo

$$\begin{aligned} \vec{X}^H A\vec{X} &= (U\vec{Z})^H A(U\vec{Z}) && \text{puesto que } \vec{X} = U\vec{Z} \\ &= (\vec{Z}^H\ U^H)A(U\vec{Z}) \\ &= \vec{Z}^H(U^H AU)\vec{Z} = \vec{Z}^H D\vec{Z} && \text{puesto que } U \text{ diagonaliza a } A. \end{aligned}$$

■

A continuación, se presentan los pasos a seguir para determinar la diagonalización de una forma hermitiana mediante este método:

Procedimiento para diagonalizar una forma hermítica

i) Halle la matriz de coeficientes hermitiana A asociada a $F(\vec{X})$.

ii) Encuentre los valores propios (no necesariamente distintos), $\lambda_1, \lambda_2, \ldots, \lambda_n$ de A.

iii) Encuentre una base ortonormal para $\mathbb{C}^n$ formada por los vectores propios normalizados de A.

iv) Forme la matriz U cuyas columnas sean los vectores de la base hallada en el paso *iii)* en el orden correspondiente al listado de los valores propios en el paso *ii)*. La transformación $\vec{X} = U\vec{Z}$ es una *rotación* si $\| \det(U)\| = 1$.

v) Transforme a $F(\vec{X})$ en $\lambda_1|z_1|^2 + \lambda_2|z_2|^2 + \ldots + \lambda_n|z_n|^2$.

Ejemplo 6.5 *Considere la ecuación cuadrática compleja dada en el Ejemplo 6.4. Determine la "superficie" cuadrática obtenida al eliminar los términos de productos cruzados.*

Solución.

En el Ejemplo 4.6, se obtuvo que la matriz A asociada a la forma cuadrática compleja era diagonalizable mediante la matriz unitaria

$$U = \begin{bmatrix} 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ i/\sqrt{3} & 1/\sqrt{2} & -i/\sqrt{6} \\ -1/\sqrt{3} & -i/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}.$$

Luego,

$$U^H A U = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Por consiguiente, (6.16) se puede escribir en términos de las nuevas variables z_1, z_2, z_3 como $\vec{Z}^H D \vec{Z}$, es decir,

$$3|z_1|^2 - 2|z_2|^2 = 9, \tag{6.18}$$

donde

$$\vec{Z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = U^H \vec{X} = \begin{bmatrix} 1/\sqrt{3} & -i/\sqrt{3} & -1/\sqrt{3} \\ 0 & 1/\sqrt{2} & i/\sqrt{2} \\ 2/\sqrt{6} & i/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \tag{6.19}$$

y reescribiendo (6.18), se obtiene

$$\frac{1}{3}|z_1|^2 - \frac{1}{9/2}|z_2|^2 = 1,$$

la cual corresponde a la ecuación de una "hipérbola" sobre los ejes z_1 y z_2.

Definición 6.4 Forma polar de una forma hermitiana

Dada F una forma cuadrática compleja, se puede obtener una forma hermítica g de F de acuerdo con la siguiente identidad llamada la forma polar de g:

$$g(\vec{u}, \vec{v}) = \frac{1}{4}\big[F(\vec{u} + \vec{v}) - F(\vec{u} - \vec{v})\big] + \frac{i}{4}\big[F(\vec{u} + i\vec{v}) - F(\vec{u} - i\vec{v})\big]. \tag{6.20}$$

6.4 Clasificación de formas cuadráticas complejas

Definición 6.5 *Una forma cuadrática compleja* $F(\vec{X}) = \vec{X}^H A\vec{X}$ *asociada a una matriz* $A \neq \mathrm{O}$*, es*

1. **Definida positiva** *si* $F(\vec{X}) > 0$ *para todo* $\vec{X} \neq \vec{0}$ *en* $\mathbb{C}^n$.

2. **Definida negativa** *si* $F(\vec{X}) < 0$ *para todo* $\vec{X} \neq \vec{0}$ *en* $\mathbb{C}^n$.

3. **Indefinida** *si* $F(\vec{X})$ *asume ambos valores positivos y negativos.*

4. **Semidefinida positiva** *si* $F(\vec{X}) \geq 0$ *para todo* $\vec{X} \in \mathbb{C}^n$.

5. **Semidefinida negativa** *si* $F(\vec{X}) \leq 0$ *para todo* $\vec{X} \in \mathbb{C}^n$.

La matriz hermitiana asociada A *se denomina definida positiva, semidefinida positiva, definida negativa, semidefinida negativa o indefinida según sea la forma cuadrática compleja* $F(\vec{X})$ *que define.*

Ejemplo 6.6 *Verifique si la forma hermítica dada en el Ejemplo 6.2 es definida positiva.*

Solución.

La forma $F(\vec{v})$ dada en el Ejemplo 6.2 es definida positiva ya que para todo $\vec{v} \neq \vec{0}$:

$$F(\vec{v}) = \vec{v}^H \vec{v} = \sum_{i=1}^{n} \overline{z_i} z_i = \sum_{i=1}^{n} |z_i|^2 > 0,$$

donde $\vec{v}^H = \overline{(z_1, z_2, \ldots, z_n)^t}$ y $z_i \in \mathbb{C}$.

6.5 Orden parcial entre matrices

Dadas dos matrices, además de combinarlas haciendo operaciones entre ellas (suma, resta, multiplicación), las podemos comparar para ordenarlas o clasificarlas. Una comparación que surgió en secciones anteriores fue ver si eran semejantes. En esta sección se hablará de un orden "parcial" entre matrices semidefinidas positivas.

Definición 6.6 Orden entre matrices

Sean A y B matrices hermitianas de tamaño $n \times n$. Se escribe $A \succcurlyeq B$ si la matriz $A - B$ es semidefinida positiva. Similarmente, $A \succ B$ significa que la matriz $A - B$ es definida positiva.

Teorema 6.8 *Si A y B son matrices hermitianas de tamaño $n \times n$, entonces:*

$$A \succcurlyeq B \qquad \textit{implica que} \qquad T^H A T \succcurlyeq T^H B T$$

para toda $T \in M_{nn}$.

Demostración.
Si $A - B$ es semidefinida positiva, entonces $\vec{Y}^H \big(A - B\big)\vec{Y} \geq 0$ para todo $\vec{Y} \in \mathbb{C}^n$. Así,

$$\vec{X}^H \big(T^H A T - T^H B T\big)\vec{X} = \big(T\vec{X}\big)^H \big(A - B\big)\big(T\vec{X}\big) \geq 0$$

para todo $\vec{X} \in \mathbb{C}^n$, lo cual, a su vez, significa que $T^H A T \succcurlyeq T^H B T$. ■

Corolario 6.8.1 *Si A y B son matrices de tamaño $n \times n$ definidas positivas, entonces:*

a) Si $A \succcurlyeq B$, entonces $\lambda_k(A) \geq \lambda_k(B)$ para todo $k = 1, 2, \ldots, n$ si los respectivos valores propios se colocan en el mismo orden (creciente o decreciente).

b) Si $A \succcurlyeq B$, *entonces* $\det A \geq \det B$ *y* $\operatorname{tr}(A) \geq \operatorname{tr}(B)$.

c) $A \succ B$ *si y solo si* $B^{-1} \succ A^{-1}$.

Demostración.
Queda como ejercicio para el lector. ■

Ejemplo 6.7 *Dadas las siguientes matrices:*

$$A = \begin{bmatrix} 7 & 1 & 2 \\ 1 & 7 & 2 \\ 2 & 2 & 10 \end{bmatrix} \qquad y \qquad B = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix},$$

muestre que A *es más positiva que* B.

Solución.
Primero, se obtiene la matriz $C = A - B$:

$$C = \begin{bmatrix} 5 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 8 \end{bmatrix}.$$

Los valores propios de esta matriz son

$$\lambda_1 = 5, \qquad \lambda_2 = 5 + \sqrt{11}, \qquad \lambda_3 = 5 - \sqrt{11},$$

como todos los valores propios son positivos, entonces C es definida positiva y, por lo tanto, A es más positiva que B.

Ejercicios 6.1

1. Reduzca las siguientes formas complejas a una forma diagonal:

   a) $2|x_1|^2 + (1-i)\overline{x_1}x_2 + (1+i)\overline{x_2}x_1 + 3|x_2|^2 = 4.$

   b) $|x_1|^2 - i\overline{x_1}x_2 + i\overline{x_2}x_1 + \overline{x_1}x_3 + \overline{x_3}x_1 - i\overline{x_2}x_3 + i\overline{x_3}x_2 - 2|x_3|^2 = 5.$

2. Considere las matrices asociadas a cada una de las formas cuadráticas complejas del Ejercicio 1. Determine qué tipo de matrices son (definida positiva o definida negativa).

3. Sea A una matriz hermítica cuadrada de tamaño 3×3. Supóngase que $\Delta_1 > 0$, $\Delta_2 < 0$ y $\Delta_3 > 0$. Muestre que A tiene un valor propio positivo y dos negativos.

4. Sea A cualquier matriz compleja no singular. Muestre que $B = A^H A$ es hermítica y definida positiva.

5. Muestre que si A es una matriz hermítica cuadrada de tamaño $n \times n$ definida positiva con valores propios $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$ y B es una submatriz principal de A de tamaño $k \times k$, entonces

$$\prod_{j=1}^{k} \lambda_{n-j+1} \leq \det B \leq \prod_{j=1}^{k} \lambda_j.$$

6. Sean A y B matrices hermíticas cuadradas de tamaño $n \times n$ con valores propios $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ y $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$, respectivamente. Sean $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ los valores propios de $A + B$. Para $1 \leq k \leq n$, demuestre que

$$\text{máx}\big\{\lambda_k + \mu_n, \lambda_n + \mu_k\big\} \leq \sigma_k \leq \text{máx}\big\{\lambda_k + \mu_1, \lambda_1 + \mu_k\big\}.$$

Capítulo 7

Normas matriciales

En este capítulo se intenta medir la sensibilidad o la "vulnerabilidad" de la solución de un sistema no singular de ecuaciones lineales $A\vec{X} = \vec{b}$. En otras palabras, se quiere medir qué tan grande es el efecto en $\vec{X} = A^{-1}\vec{b}$ si se cambian ligeramente las componentes de A y $\vec{b}$. Es decir, se debe encontrar una manera de medir el cambio ΔA y definir la "longitud" de una matriz, pues para vectores ya sabemos cómo obtener su longitud y ahora necesitamos un concepto análogo para matrices.

7.1 Definición y resultados básicos

Definición 7.1 Norma de una matriz

Sea $\mathcal{M}_{nn}$ el espacio de las matrices de tamaño $n \times n$ con componentes reales (complejas). Una norma de matriz $\|\cdot\|$ de $\mathcal{M}_{nn}$ en $\mathbb{R}$ es una función que satisface para toda $A,\ B \in \mathcal{M}_{nn}$ los cinco axiomas siguientes:

(1) $\|A\| \geq 0$	*No negativa*
(2) $\|A\| = 0$ *si y solo si* $A = 0$	*Positiva*
(3) $\|cA\| = \|c\|\|A\|$ *para todo escalar* c	*Homogénea*

(4) $\|A+B\| \leq \|A\| + \|B\|$ *Desigualdad triangular*

(5) $\|AB\| \leq \|A\|\ \|B\|$ *Submultiplicativa.*

Teorema 7.1 *Sea $\|\cdot\|$ cualquier norma matricial, entonces:*

1. $\|I_n\| \geq 1$, *donde I_n es la matriz identidad de tamaño $n \times n$.*
2. $\|A^{-1}\| \geq \dfrac{\|I_n\|}{\|A\|}$ *para cualquier matriz $A \in \mathcal{M}_{nn}$ no singular.*
3. $\left\|A^k\right\| \leq \|A\|^k$ *para cualquier matriz $A \in \mathcal{M}_{nn}$ y todo $k \geq 2$.*

Demostración.

1. Queda como ejercicio para el lector.
2. Puesto que $AA^{-1} = I_n$, entonces

$$\|I_n\| = \|AA^{-1}\| \leq \|A\|\ \|A^{-1}\| \qquad \text{por propiedad (5).}$$

   Pero como $\|A\| > 0$, se tiene que

$$\|A^{-1}\| \geq \frac{\|I_n\|}{\|A\|}.$$

3. La demostración es por inducción sobre k. El resultado es trivial para $k = 2$, puesto que por la propiedad submultiplicativa

$$\|A^2\| \leq \|A\|\ \|A\| = \|A\|^2.$$

   Supongamos que se ha demostrado para cualquier $k = m$, es decir

$$\|A^m\| \leq \|A\|^m.$$

   Luego, $\|A^{m+1}\| = \|A^m A\|$ y, por la propiedad submultiplicativa, se tiene

$$\|A^m A\| \leq \|A^m\|\|A\| \leq \|A\|^m \|A\| = \|A\|^{m+1}.$$

■

7.2 Tipos de normas matriciales

Las normas que se consideran en esta sección son algunas de las normas que se pueden emplear en el espacio de matrices $\mathcal{M}_{nn}$ análogas a las normas de los espacios vectoriales.

Teorema 7.2 Norma L_1

Dada la matriz $A = [a_{ij}]$, la función $\|\cdot\| : \mathcal{M}_{nn} \to \mathbb{R}$ definida por

$$\|A\|_1 = \sum_{i,j=1}^{n} |a_{ij}|$$

es una norma de matriz.

Demostración.

Los axiomas de $(1) - (3)$ se satisfacen fácilmente de la definición de valor absoluto, se demostrará, por lo tanto, que se cumplen los axiomas (4) y (5).

(4) Para la desigualdad triangular se tiene

$$\begin{aligned}\|A+B\|_1 &= \sum_{i,j=1}^{n} |a_{ij}+b_{ij}| \leq \sum_{i,j=1}^{n} (|a_{ij}|+|b_{ij}|) \\ &= \sum_{i,j=1}^{n} |a_{ij}| + \sum_{i,j=1}^{n} |b_{ij}| = \|A\|_1 + \|B\|_1.\end{aligned}$$

Por lo tanto, se cumple el axioma (4).

(5) Puesto que $AB = C = [c_{ij}] = \sum_{k=1}^{n} a_{ik}\ b_{kj}$, entonces

$$\begin{aligned}\|AB\|_1 &= \sum_{i,j=1}^{n} \left| \sum_{k=1}^{n} a_{ik}\ b_{kj} \right| \leq \sum_{i,j,k=1}^{n} |a_{ik}\ b_{jk}| \leq \sum_{i,j,k,m=1}^{n} |a_{ik}\ b_{mj}| \\ &= \left[\sum_{i,k=1}^{n} |a_{ik}| \right] \left[\sum_{j,m=1}^{n} |b_{mj}| \right] = \|A\|_1\ \|B\|_1.\end{aligned}$$

En la verificación de este axioma, la primera desigualdad se obtiene de la generalización de la desigualdad triangular y la segunda, de los términos adicionales a la suma.

Por consiguiente, $\|A\|_1$ sí es una norma matricial. ■

Ejemplo 7.1 Norma euclideana (L_2)

Determine si la norma L_2 definida por

$$\|A\|_2 = \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}}$$

es una norma de matriz.

Solución.

Fácilmente se puede probar que los axiomas de $(1)-(3)$ se satisfacen. Por lo tanto, veamos si se cumplen los axiomas (4) y (5):

$$\begin{aligned}
(4)\ \|A+B\|_2^2 &= \sum_{i,j=1}^{n} |a_{ij}+b_{ij}|^2 \leq \sum_{i,j=1}^{n} \left(|a_{ij}|^2 + 2|a_{ij}|\ |b_{ij}| + |b_{ij}|^2 \right) \\
&= \sum_{i,j=1}^{n} |a_{ij}|^2 + 2 \sum_{i,j=1}^{n} |a_{ij}\ b_{ij}| + \sum_{i,j=1}^{n} |b_{ij}|^2 \\
&\leq \left[\left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}} + \left(\sum_{i,j=1}^{n} |b_{ij}|^2 \right)^{\frac{1}{2}} \right]^2 \\
&= \left(\|A\|_2 + \|B\|_2 \right)^2 .
\end{aligned}$$

Luego, se cumple el axioma (4).

$$\begin{aligned}
(5)\ \|AB\|_2^2 &= \sum_{i,j=1}^{n} \left| \sum_{k=1}^{n} a_{ik}\ b_{kj} \right|^2 \leq \sum_{i,j=1}^{n} \left[\left(\sum_{k=1}^{n} |a_{ik}|^2 \right) \left(\sum_{m=1}^{n} |b_{mj}|^2 \right) \right] \\
&= \left(\sum_{i,k=1}^{n} |a_{ik}|^2 \right) \left(\sum_{j,m=1}^{n} |b_{mj}|^2 \right) = \|A\|_2^2\ \|B\|_2^2 .
\end{aligned}$$

Esta desigualdad es justo la *desigualdad de Cauchy-Schwarz*. Por consiguiente, $\|\cdot\|_2$ sí es una norma.

Ejemplo 7.2 Norma L_∞

Determine si la función $\|\cdot\| : \mathcal{M}_{nn} \to \mathbb{R}$ definida por

$$\|A\|_\infty = n \max_{1 \leq i,j \leq n} |a_{ij}|$$

es una norma de matriz.

Solución.

Los axiomas de (1) − (3) se satisfacen fácilmente de la definición de máximo. Se probará si se cumplen los axiomas (4) y (5):

$$\begin{aligned}(4)\ \|A+B\|_\infty &= n \max_{1\le i,j\le n} |a_{ij}+b_{ij}| \le n \max_{1\le i,j\le n} (|a_{ij}|+|b_{ij}|)\\ &= n \max_{1\le i,j\le n} |a_{ij}| + n \max_{1\le i,j\le n} |b_{ij}| = \|A\|_\infty + \|B\|_\infty.\end{aligned}$$

Por lo tanto, se cumple el axioma (4).

$$\begin{aligned}(5)\ \|AB\|_\infty &= n \max_{1\le i,j\le n} \left| \sum_{k=1}^{n} a_{ik}\ b_{kj} \right| \le n \max_{1\le i,j\le n} \sum_{k=1}^{n} |a_{ik}\ b_{kj}|\\ &\le n \max_{1\le i,j\le n} \sum_{k=1}^{n} \frac{\|A\|_\infty}{n}\ \frac{\|B\|_\infty}{n} = \|A\|_\infty\ \|B\|_\infty.\end{aligned}$$

Por consiguiente, $\|A\|_\infty$ sí es una norma.

Definición 7.2 Norma matricial inducida

Sea $\|\cdot\|$ una norma vectorial sobre $\mathbb{C}^n$. Se define $\|\cdot\|_{in}$ sobre $\mathcal{M}_{nn}$ por

$$\|A\|_{in} = \max_{\|\vec{x}\|=1} \|A\vec{x}\| = \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|}{\|\vec{x}\|}. \tag{7.1}$$

Las letras "in" en la norma es la abreviación de la frase "norma inducida".

Teorema 7.3 Norma espectral

La norma espectral $\|\cdot\|_S$ se define sobre $\mathcal{M}_{nn}$ por

$$\|A\|_S = \sigma_1 = \max\{\sigma_i : \sigma_i \text{ es un valor singular de } A\}. \tag{7.2}$$

Demostración.

De la ecuación (7.1), se tiene que

$$\|A\|_{in}^2 = \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} = \max_{\vec{x}\neq\vec{0}} \left\{ \frac{(A\vec{x})^H (A\vec{x})}{\vec{x}^H\vec{x}} \right\}.$$

Luego, si $A^H A\vec{x} = \sigma^2\vec{x}$, se obtiene

$$\|A\|_{in}^2 = \max_{\vec{x}\neq\vec{0}}\left\{\frac{\vec{x}^H(A^HA\vec{x})}{\vec{x}^H\vec{x}}\right\} = \max_{\vec{x}\neq\vec{0}}\left\{\sigma^2\frac{\vec{x}^H\vec{x}}{\vec{x}^H\vec{x}}\right\} = \sigma^2_{\text{máx}}$$

como A^HA es una matriz hermitiana, sus valores propios son reales. ■

Definición 7.3 Radio espectral

El radio espectral $r(A)$ de una matriz $A \in \mathcal{M}_{nn}$ es definido por la cantidad:

$$r(A) = \max\{\|\lambda\| : \lambda \text{ es un valor propio de } A\},$$

donde $\|\lambda\|$ es el valor absoluto o módulo de λ.

Ejemplo 7.3 *Obtenga la norma espectral y el radio espectral de la matriz:*

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

Solución.

Como la matriz A es simétrica, sus valores singulares y sus valores propios son iguales, es decir, $\sigma_1 = \lambda_1 = 4$ y $\sigma_2 = \lambda_2 = 2$. Por lo tanto,

$$\|A\|_S = 4 \qquad \text{y} \qquad r(A) = 4.$$

Teorema 7.4 *Sea $A \in \mathcal{M}_{nn}$ y $\|\cdot\|$ cualquier norma de matriz, entonces:*

$$r(A) \leq \|A\|.$$

Demostración.

Supóngase que $A\vec{x} = \lambda\vec{x}$, $\vec{x} \neq \vec{0}$ y que $|\lambda| = r(A)$. Sea $X \in \mathcal{M}_{nn}$ la matriz cuyas columnas son todas iguales a $\vec{x}$, entonces $AX = \lambda X$. Luego, si $\|\cdot\|$ es cualquier norma de matriz, se tiene que

$$\begin{aligned} \|AX\| &\leq \|A\|\ \|X\| \\ \|\lambda X\| &\leq \|A\|\ \|X\| \\ \|\lambda\|\ \|X\| &\leq \|A\|\ \|X\|. \end{aligned}$$

Por consiguiente, $\|\lambda\| = r(A) \leq \|A\|$. ■

Ejercicios 7.1

1. Calcule para cada una de las siguientes matrices la norma espectral y el radio espectral:

$$a.\begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}. \qquad b.\begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix}. \qquad c.\begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}.$$

$$d.\begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -1 & 1 & 1 \end{bmatrix}. \qquad e.\begin{bmatrix} 3 & -1 & 4 \\ -4 & 1 & -5 \\ 5 & 3 & 2 \end{bmatrix}.$$

2. Muestre que A y A^t tienen el mismo radio espectral y la misma norma espectral.

3. Si A es una matriz simétrica de tamaño $n \times n$, muestre que su norma espectral coincide con su radio espectral.

4. Si A es una matriz hermitiana de tamaño $n \times n$, muestre que la norma espectral y el radio espectral son iguales.

Teorema 7.5 Lema de Banach

Sea A una matriz real de tamaño $n \times n$ y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. Suponiendo que $\|A\| < 1$, entonces $I_n - A$ es no singular y

$$\frac{1}{1+\|A\|} \leq \left\| \left(I_n - A\right)^{-1} \right\| \leq \frac{1}{1-\|A\|}.$$

Demostración.

La matriz $I_n - A$ es no singular si y solo si la única solución del sistema homogéneo $\left(I_n - A\right)\vec{x} = \vec{0}$ es $\vec{x} = \vec{0}$. Suponga entonces que $\left(I_n - A\right)\vec{x} = \vec{0}$ de modo que $\vec{x} = A\vec{x}$. Entonces, se tiene que

$$\|\vec{x}\| = \|A\vec{x}\| \leq \|A\|\|\vec{x}\|.$$

Pero como $\|A\| < 1$, entonces hay una contradicción a menos que $\vec{x} = \vec{0}$, como se ha tratado probar. Así que $\big(I_n - A\big)^{-1}$ existe, se denotará con R. Luego,

$$I_n = R\big(I_n - A\big) = R - RA. \tag{7.3}$$

Por lo tanto,

$$1 = \|I_n\| = \big\|R\big(I_n - A\big)\big\| \leq \|R\|\,\|I_n + (-A)\| \leq \|R\|\,(1 + \|A\|)\,,$$

de manera que $\|R\| \geq 1/\,(1 + \|A\|)$ como se afirmó. De la expresión (7.3), $R = I_n + RA$, así que

$$\|R\| = \|I_n + RA\| \leq 1 + \|RA\| \leq 1 + \|R\|\|A\|.$$

Por consiguiente, $\|R\| \leq \big(1 - \|A\|\big)^{-1}$, lo cual completa la prueba. ■

Ejemplo 7.4 *Para la siguiente matriz:*

$$A = \frac{1}{10}\begin{bmatrix} 11 & -6 \\ 8 & 9 \end{bmatrix}$$

determine las cotas superior e inferior del lema de Banach.

Solución.

La matriz A se puede escribir como $A = I - B$, en donde

$$B = \frac{1}{10}\begin{bmatrix} -1 & 6 \\ -8 & 1 \end{bmatrix}.$$

Como se puede emplear cualquier norma matricial, usando la norma espectral, se tiene que

$$\|B\|_S = \frac{7 + \sqrt{2}}{10} < 1.$$

El lema de Banach dice que $A = I - B$ es no singular, y por lo tanto

$$\frac{10}{17 + \sqrt{2}} \leq \big\|A^{-1}\big\|_S \leq \frac{10}{3 - \sqrt{2}}.$$

El lema de Banach dice que matrices suficientemente "cercanas" a I_n son no singulares. El teorema siguiente es una generalización de este hecho.

Teorema 7.6 Inversas perturbadas

Sean A y B matrices de tamaño $n \times n$ siendo A no singular y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. Defínase $\alpha = \left\|A^{-1}B\right\|$ o $\alpha = \left\|BA^{-1}\right\|$. Si $\alpha < 1$ (es decir en especial si $\|B\| < 1/\left\|A^{-1}\right\|$), entonces $A - B$ también es no singular y

$$\frac{\left\|A^{-1}\right\|}{1+\alpha} \leq \left\|(A-B)^{-1}\right\| \leq \frac{\left\|A^{-1}\right\|}{1-\alpha}.$$

Demostración.

Supongamos que $\left\|A^{-1}B\right\| < 1$. El otro caso es semejante. Como A^{-1} existe, se puede escribir $A - B$ como

$$A\left(I_n - A^{-1}B\right) = A\big(I_n - R\big),$$

donde $R = A^{-1}B$. Por hipótesis, $\|R\| = \alpha < 1$, de modo que al aplicar el lema de Banach, se obtiene que $I_n - R$ es no singular, como lo es A. Luego,

$$A\big(I_n - R\big) = A - B \tag{7.4}$$

es también no singular y

$$\big(A-B\big)^{-1} = \big[A\big(I_n - R\big)\big]^{-1} = \big(I_n - R\big)^{-1}A^{-1}.$$

Por lo tanto,

$$\left\|\big(A-B\big)^{-1}\right\| \leq \left\|\big(I_n - R\big)^{-1}\right\| \left\|A^{-1}\right\| \leq \frac{\left\|A^{-1}\right\|}{1-\alpha}.$$

Por el lema de Banach, esta es la cota superior que se deseaba. Para obtener la cota inferior, se reescribe (7.4) como

$$A^{-1} = (I_n - R)\left(A-B\right)^{-1},$$

de lo cual se deduce que

$$\left\|A^{-1}\right\| \leq \|I_n - R\| \left\|(A-B)^{-1}\right\| \leq (1+\alpha)\left\|(A-B)^{-1}\right\|.$$

Al dividir por $(1+\alpha)$, se obtiene la cota inferior que se buscaba. ■

7.3 Condición de sistemas de ecuaciones lineales

El concepto de *condición* es importante en todas las matemáticas aplicadas. Si "pequeños cambios en los datos" de un problema producen cambios razonablemente pequeños en su solución, se dice que el problema está *bien planteado*. Si "pequeños cambios en los datos" de algún problema ocasionan cambios inaceptablemente grandes en la solución, se dice que el problema está *mal planteado*. La razón de la importancia de este concepto debería ser evidente: en los problemas aplicados, casi siempre los datos son inexactos por errores de medición y de modelamiento, y es crucial conocer los efectos que tienen las inexactitudes en los datos sobre la solución del problema.

Definición 7.4 Sistema de ecuaciones de mal comportamiento

Un sistema de ecuaciones lineales

$$A\vec{X} = \vec{b} \tag{7.5}$$

con A una matriz de tamaño $n \times n$, $\vec{X} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^n$, se dice que es un sistema de mal comportamiento si las n columnas de la matriz son casi linealmente dependientes o, en otras palabras, si la matriz de los coeficientes es casi singular. Esto significa que un cambio pequeño en algunos elementos de A produce una matriz singular.

Definición 7.5 Número de condición

Sea A una matriz no singular real de tamaño $n \times n$, el número de condición se define como

$$\kappa\big(A\big) = \|A\|_S \|A^{-1}\|_S = \frac{\sigma_{max}}{\sigma_{min}}, \tag{7.6}$$

donde, σ_{max} y σ_{min} son los valores singulares más grande y más pequeño, respectivamente, asociados a A.

Teorema 7.7 *Sea A no singular y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. La sensibilidad de la solución de un sistema de ecuaciones lineales*

$$A\vec{x} = \vec{b}$$

con respecto a la perturbación ΔA en A se relaciona directamente con el número de condición. En otras palabras, si

$$\vec{y} \qquad \text{resuelve a} \qquad (A + \Delta A)\,\vec{y} = \vec{b}.$$

Entonces, el cambio en la solución satisface que

$$\frac{\|\vec{y} - \vec{x}\|}{\|\vec{y}\|} \leq \alpha\kappa(A),$$

en donde $\alpha = \|\Delta A\| / \|A\|$ es el error relativo en A.

Demostración.

Puesto que la solución del sistema perturbado es $\vec{y}$, entonces

$$\begin{aligned}
(A + \Delta A)\,\vec{y} - \vec{b} &= \vec{0} \\
A\vec{y} + \Delta A\vec{y} - A\vec{x} &= \vec{0} \\
A\,(\vec{y} - \vec{x}) &= -\Delta A\vec{y} \\
\vec{y} - \vec{x} &= A^{-1}\,(-\Delta A\vec{y})\,.
\end{aligned}$$

Luego, si $\|\cdot\|$ es cualquier norma matricial, se tiene que

$$\begin{aligned}
\|\vec{y} - \vec{x}\| &= \left\|-A^{-1}\Delta A\vec{y}\right\| \leq \left\|A^{-1}\Delta A\right\| \|\vec{y}\| \\
&\leq \left\|A^{-1}\right\| \|\Delta A\| \|\vec{y}\|\,,
\end{aligned}$$

como se quería. ■

Ejemplo 7.5 *Resuelva el sistema de ecuaciones:*

$$(1 + \epsilon)\,x_1 + x_2 = 2, \qquad\qquad x_1 + x_2 = 1,$$

con $\epsilon \in \mathbb{R}$ y en las fórmulas que obtenga para x_1 y x_2, sustituya:

$$(i)\ \epsilon = 0.01, 0.02 \qquad y \qquad (ii)\ \epsilon = 2.01, 2.04.$$

Compare los cambios en porcentaje del coeficiente de x_1 en la primera ecuación, en los casos (i) y (ii), con los cambios en porcentaje de la correspondiente solución de x_1.

Solución.

Aplicando el método de eliminación de Gauss, se obtiene

$$\left[\begin{array}{ccc} 1+\epsilon & 1 & 2 \\ 1 & 1 & 1 \end{array}\right] \underset{F_1 - F_2}{\sim} \left[\begin{array}{ccc} \epsilon & 0 & 1 \\ 1 & 1 & 1 \end{array}\right].$$

Esto conduce a $x_1 = \frac{1}{\epsilon}$ y $x_2 = 1 - \frac{1}{\epsilon}$, $(\epsilon \neq 0)$

	(i)		(ii)	
ϵ	0.01	0.02	2.01	2.04
Solución	$(100, -99)$	$(50, -49)$	$\left(\frac{100}{201}, \frac{101}{201}\right)$	$\left(\frac{25}{51}, \frac{26}{51}\right)$

El cambio en porcentaje del coeficiente de x_1 es $1\,\%$ con una cifra decimal en cada caso.
La solución de x_1 en (i) varía en un $50\,\%$ y en (ii) en un $1.47\,\%$. Esto indica que cuando ϵ es pequeño, las "ecuaciones son de mal comportamiento".

Un modo sencillo de probar si un sistema de ecuaciones es de mal comportamiento consiste, precisamente, en proceder como lo hicimos en el ejemplo anterior, esto es, efectuar un pequeño cambio en algunos coeficientes para ver qué efectos se producen en la solución, pero esto es difícil de hacer cuando se trata de un sistema de ecuaciones muy grande. Existe un método que nos da una indicación de cuándo se presenta el mal comportamiento, usando la Definición 7.5.

El número de condición nos da una regla práctica para determinar si un sistema de ecuaciones es de mal comportamiento:

Si	$0 \leq \kappa(A) \leq 100$	siempre el sistema es bien condicionado,
	$100 < \kappa(A) \leq 1000$	a veces el sistema es mal condicionado,
	$1000 < \kappa(A)$	siempre el sistema es mal condicionado.

Ejemplo 7.6 *Determine el número de condición de la matriz asociada al sistema de ecuaciones del ejemplo anterior.*

Solución.

Como la matriz asociada al sistema es simétrica, su norma espectral coincide con su radio espectral; luego, se necesitan los valores propios de la matriz A. En este caso, el polinomio característico de A es

$$p_A(\lambda) = \lambda^2 - (2+\epsilon)\,\lambda + \epsilon,$$

de donde los valores propios son

$$\lambda_1 = 1 + \frac{1}{2}\epsilon + \frac{1}{2}\sqrt{4+\epsilon^2} \qquad \text{y} \qquad \lambda_2 = 1 + \frac{1}{2}\epsilon - \frac{1}{2}\sqrt{4+\epsilon^2}.$$

Luego, el número de condición de la matriz A es

$$\kappa(A) = \frac{1}{\epsilon}\left(1 + \frac{1}{2}\epsilon + \sqrt{1 + \left(\frac{1}{2}\epsilon\right)^2}\right)^2.$$

Para los valores de ϵ dados en el Ejemplo 7.5, se tiene que

	(i)		(ii)	
ϵ	0.01	0.02	2.01	2.04
$\kappa(A)$	402.01	202.02	5.8285	5.8292

Ejemplo 7.7 *Suponga que la matriz de covarianza de un experimento con tres variables x_1, x_2 y x_3 es:*

$$S = \frac{1}{10}\begin{bmatrix} 4 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Encuentre el número de condición.

Solución.

Como S es simétrica, entonces su norma espectral coincide con su radio espectral; por lo tanto, se necesitan los valores propios de la matriz S. En este caso, el polinomio característico de S es

$$p_S(\lambda) = -\lambda^3 + \lambda^2 - \frac{27}{100}\lambda + \frac{9}{500},$$

de donde los valores propios son

$$\lambda_1 = \frac{3}{5}, \qquad \lambda_2 = \frac{3}{10} \qquad \text{y} \qquad \lambda_3 = \frac{1}{10}.$$

Luego, el número de condición es

$$\kappa(A) = \frac{\lambda_1}{\lambda_3} = \frac{3/5}{1/10} = 6.$$

Como $\kappa(A)$ es pequeño (< 100), significa que cambios pequeños en los datos producen cambios razonablemente pequeños en la estimación de la matriz de covarianza.

Teorema 7.8 *Sea A no singular y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. La sensibilidad de la solución de un sistema de ecuaciones lineales*

$$A\vec{x} = \vec{b}$$

con respecto a la perturbación $\Delta\vec{b}$ en $\vec{b}$ se relaciona directamente con el número de condición. En otras palabras, si

$$A\vec{y} = \vec{b} + \Delta\vec{b} \qquad \text{con} \qquad \frac{\|\Delta\vec{b}\|}{\|\vec{b}\|} \leq \alpha,$$

entonces el cambio en la solución satisface que

$$\frac{\|\vec{y} - \vec{x}\|}{\|\vec{x}\|} \leq \alpha\kappa(A),$$

en donde $\alpha = \|\Delta A\| / \|A\|$ es el error relativo en A.

Demostración.

Puesto que la solución del sistema perturbado es $\vec{y}$, entonces

$$\begin{aligned} A\vec{y} &= \vec{b} + \Delta\vec{b} \\ A\vec{y} - A\vec{x} - \Delta\vec{b} &= \vec{0} \\ A\left(\vec{y} - \vec{x}\right) &= \Delta\vec{b} \\ \vec{y} - \vec{x} &= A^{-1}\Delta\vec{b}. \end{aligned}$$

Luego, si $\|\cdot\|$ es cualquier norma matricial, se tiene que

$$\begin{aligned} \|\vec{y} - \vec{x}\| &= \left\|A^{-1}\Delta\vec{b}\right\| \leq \left\|A^{-1}\right\| \left\|\Delta\vec{b}\right\| \\ &\leq \alpha \left\|A^{-1}\right\| \left\|\vec{b}\right\|. \end{aligned}$$

Pero como $\vec{b} = A\vec{x}$, se tiene que $\left\|\vec{b}\right\| \leq \|A\| \|\vec{x}\|$. Entonces,

$$\|\vec{y} - \vec{x}\| \leq \alpha\kappa(A) \|\vec{x}\|.$$

Así, se completa la prueba. ■

Definición 7.6 Índice de condición

Sea A una matriz real de tamaño $n \times n$, el índice de condición se define como

$$IC\big(A\big) = \sqrt{\kappa\big(A\big)}, \tag{7.7}$$

donde $\kappa\big(A\big)$ es el número de condición de A.

Teorema 7.9 *Sea A no singular y sea $\|\cdot\|$ una norma matricial sobre $\mathcal{M}_{nn}$. Suponga que $\vec{x}$ resuelve a $A\vec{x} = \vec{b}$ mientras que*

$$\vec{y} = \vec{x} + \Delta\vec{x} \qquad \textit{resuelve a} \qquad \left(A + \Delta A\right)\vec{y} = \vec{b} + \Delta\vec{b},$$

para ciertas perturbaciones ΔA y $\Delta\vec{x}$ en los datos. Suponga que la perturbación ΔA es lo suficientemente pequeña como para que $\alpha < 1$, en

donde $\alpha = \left\|(\Delta A)\, A^{-1}\right\|$ *o* $\alpha = \left\|A^{-1}\,(\Delta A)\right\|$. *Entonces, el cambio* $\Delta\vec{x}$ *en la solución satisface*

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \frac{\kappa(A)}{1-\alpha}\left(\frac{\|\Delta\vec{b}\|}{\|\vec{b}\|} + \frac{\|\Delta A\|}{\|A\|}\right),$$

en donde $\kappa(A)$ *es el número de condición de* A.

Demostración.

Si $0 < \alpha < 1$, el Teorema 7.6 implica que $A + \Delta A$ es no singular y da una cota para la norma de su inversa. Como $A + \Delta A$ es no singular, la solución $\vec{y}$ al problema perturbado existe. De hecho, $\Delta\vec{x}$ resuelve el sistema

$$(A + \Delta A)\,\Delta\vec{x} = \vec{b} + \Delta\vec{b} - A\vec{x} - \Delta A\vec{x} = \Delta\vec{b} - \Delta A\vec{x}.$$

Denotando $B = -\Delta A$ y despejando $\Delta\vec{x}$ se tiene que

$$\Delta\vec{x} = \left(A - B\right)^{-1}\left(\Delta\vec{b} + B\vec{x}\right).$$

Al aplicar la norma a ambos lados y usando al lado izquierdo la cota superior dada en el Teorema 7.6, se obtiene que

$$\begin{aligned}(1-\alpha)\,\|\Delta\vec{x}\| &\leq \left\|A^{-1}\right\| \left\|\Delta\vec{b} + B\vec{x}\right\| \\ &\leq \left\|A^{-1}\right\| \left(\left\|\Delta\vec{b}\right\| + \|B\|\,\|\vec{x}\|\right).\end{aligned}$$

Por lo tanto, al reemplazar B se llega a

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \frac{\left\|A^{-1}\right\|}{1-\alpha}\left(\frac{\left\|\Delta\vec{b}\right\|}{\|\vec{x}\|} + \|\Delta A\|\right).$$

Puesto que $\vec{b} = A\vec{x}$, esto implica que $\left\|\vec{b}\right\| \leq \|A\|\,\|\vec{x}\|$. Al sustituir, se tiene que

$$\begin{aligned}\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} &\leq \frac{\left\|A^{-1}\right\|}{1-\alpha}\left(\frac{\left\|\Delta\vec{b}\right\|}{\left\|\vec{b}\right\|}\,\|A\| + \|\Delta A\|\right) \\ &\leq \frac{\kappa(A)}{1-\alpha}\left(\frac{\left\|\Delta\vec{b}\right\|}{\left\|\vec{b}\right\|} + \frac{\|\Delta A\|}{\|A\|}\right),\end{aligned}$$

lo cual completa la demostración. ■

Ejemplo 7.8 *Considere el sistema de ecuaciones:*

$$x_1 + 6x_2 = 0, \qquad 6x_1 + 46x_2 = 20, \tag{7.8}$$

que tiene la solución exacta $x_1 = -12$, $x_2 = 2$. *¿El sistema está bien condicionado o mal condicionado?*

Solución.

Para determinar si el sistema de ecuaciones dado en (7.8) es estable, se calcula el número de condición por medio de la ecuación dada en (7.6). Si se expresa matricialmente (7.8), se tiene que

$$\begin{bmatrix} 1 & 6 \\ 6 & 46 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 20 \end{bmatrix}. \tag{7.9}$$

Como la matriz del sistema es simétrica y definida positiva, su norma espectral coincide con su radio espectral; por lo tanto, se necesitan los valores propios de la matriz A. En este caso, el polinomio característico es

$$p_A(\lambda) = \lambda^2 - 47\lambda + 10,$$

de donde los valores propios son

$$\lambda_{max} = \frac{47}{2} + \frac{3}{2}\sqrt{241} \qquad \text{y} \qquad \lambda_{min} = \frac{47}{2} - \frac{3}{2}\sqrt{241}.$$

Luego, el número de condición es

$$\kappa(A) = \frac{\lambda_{max}}{\lambda_{min}} = \frac{1}{10}\left(\frac{47}{2} + \frac{3}{2}\sqrt{241}\right)^2 = \frac{1}{20}\left(2189 + 141\sqrt{241}\right) \approx 218.9.$$

En este caso, como $\kappa(A)$ es muy grande (> 100), se dice que el sistema no es estable.

Ejercicios 7.2

1. Encuentre para cada una de las siguientes matrices su número de condición y una matriz singular cercana:

$$a.\ \begin{bmatrix} 0.89 & 0.53 \\ 0.47 & 0.28 \end{bmatrix}. \qquad b.\ \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{k} \end{bmatrix}. \qquad c.\ \begin{bmatrix} 1.1 & 2.1 & 3.1 \\ 1.0 & -1.0 & 2.0 \\ 0.2 & 3.3 & 1.4 \end{bmatrix}.$$

2. Si A es una matriz simétrica real de tamaño $n \times n$, muestre que su número de condición es

$$\kappa(A) = \frac{|\lambda_{max}|}{|\lambda_{min}|},$$

donde $|\lambda_{max}|$ y $|\lambda_{min}|$ son los absolutos de los valores propios asociados a A más grande y más pequeño, respectivamente.

3. Sea A la matriz:

$$A = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \qquad \text{por lo tanto} \qquad A^{-1} = \begin{bmatrix} 1 & -k \\ 0 & 1 \end{bmatrix}.$$

Si se emplea la norma $\|\cdot\|_1$ o la norma $\|\cdot\|_\infty$, se obtiene que:

$$\|A\| = \|A^{-1}\| = 2 + k \qquad \text{para} \qquad k \geq 0.$$

Luego, el número de condición $\kappa(A) = (2+k)^2$ es grande para k grande. Sin embargo, si se considera el sistema de ecuaciones $A\vec{x} = \vec{b}$, siendo

$$\vec{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \text{se tiene que} \qquad \vec{x} = \begin{bmatrix} 1-k \\ 1 \end{bmatrix}.$$

Mientras que si solo se altera a $\vec{b}$ por medio de δ_1, δ_2 $(\neq 0)$ a

$$\vec{b} + \Delta\vec{b} = \begin{bmatrix} 1+\delta_1 \\ 1+\delta_2 \end{bmatrix}, \qquad \text{entonces} \qquad \Delta\vec{x} = \begin{bmatrix} \delta_1 - k\delta_2 \\ \delta_2 \end{bmatrix}.$$

Encuentre una cota para $\|\Delta\vec{x}\|/\|\vec{x}\|$ en términos de $\|\Delta\vec{b}\|/\|\vec{b}\|$ mediante la norma 1 o la norma ∞ para probar que este problema está bien condicionado, a pesar de que el número de condición de A es grande.

Capítulo 8

Matrices idempotentes y productos especiales

En este capítulo se enuncian algunos teoremas concernientes a un tipo especial de matriz, la *matriz idempotente* y algunos productos matriciales especiales. En muchas aplicaciones estadísticas[1] se incluyen este tipo de matrices y productos, por ello se dedica este capítulo de manera exclusiva al tratamiento de dichas matrices y productos.

8.1 Definición y propiedades

En el Capítulo 5, cuando trabajamos las formas cuadráticas, una de las condiciones que pusimos fue que la matriz asociada fuera simétrica. En el estudio de análisis de varianza, la matriz asociada a la forma cuadrática además de ser simétrica resulta ser idempotente (véase Ejemplo 5.5 y problema 6 de los Ejercicios 5.1).

Definición 8.1 Matriz idempotente

Una matriz cuadrada A se dice que es idempotente si cumple que

$$A = A^2.$$

[1] Véase Capítulo 10.

Teorema 8.1 *Los valores propios de una matriz simétrica e idempotente son cero o uno.*

Demostración.

Si A es simétrica, sus valores propios son reales. Sea λ un valor propio de A, entonces existe un vector $\vec{v} \neq \vec{0}$ tal que $A\vec{v} = \lambda\vec{v}$. Al premultiplicar ambos lados por A, se tiene que

$$A^2\vec{v} = \lambda A\vec{v} = \lambda^2\vec{v}.$$

Como A es idempotente, entonces $A^2\vec{v} = A\vec{v} = \lambda\vec{v}$, luego

$$\begin{aligned} \lambda\vec{v} &= \lambda^2\vec{v} \\ (\lambda^2 - \lambda)\vec{v} &= \vec{0}. \end{aligned}$$

Pero $\vec{v} \neq \vec{0}$, así que $\lambda^2 - \lambda$ debe ser cero. Por lo tanto, $\lambda = 0$ o $\lambda = 1$. ■

El recíproco del Teorema 8.1 no es cierto, véase el siguiente ejemplo.

Ejemplo 8.1 *Dada la matriz:*

$$A = \frac{1}{2}\begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & -1 \\ 2 & 1 & 1 \end{bmatrix},$$

obtenga los valores propios y verifique si $A^t = A$ *y* $A^2 = A$.

Solución.

En este caso, los valores propios de A son $\lambda_1 = 1$ de multiplicidad algebraica 2 y $\lambda_2 = 0$. Pero $A^t \neq A$ y $A^2 \neq A$, ya que

$$A^2 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{3}{2} & 1 & \frac{1}{2} \end{bmatrix}.$$

Por lo tanto, si los valores propios de una matriz son 0 y 1, no implica que la matriz sea simétrica e idempotente.

Teorema 8.2 *Si A es una matriz simétrica, idempotente y no singular, entonces $A = I_n$.*

Demostración.

Si A es idempotente, entonces $AA = A$. Multiplicando ambos lados por A^{-1}, se tiene lo que se quería demostrar. ■

Teorema 8.3 *Si A es idempotente, entonces $I_n - A$ es idempotente.*

Demostración.

Queda como ejercicio para el lector. ■

Teorema 8.4 *Sea A una matriz simétrica e idempotente de tamaño $n \times n$ con rango r, entonces existe una matriz ortogonal Q de tamaño $n \times n$ y una matriz R^* de tamaño $n \times n$ tal que*

$$A = QR^* \qquad y \qquad R^*Q = \begin{bmatrix} I_r & \vdots & \mathrm{O} \\ \dots & \cdot & \dots \\ \mathrm{O} & \vdots & \mathrm{O} \end{bmatrix},$$

donde I_r es la matriz identidad de tamaño $r \times r$.

Demostración.

Por el Teorema 3.33, la matriz A se puede factorizar como sigue

$$A = USV^t,$$

donde U y V son matrices ortogonales de tamaño $n \times n$ y S es una matriz diagonal de tamaño $n \times n$ con r elementos iguales a uno y los elementos restantes $n - r$ de la diagonal iguales a cero.

Puesto que $A^2 = A$, se tiene que

$$USV^tUSV^t = USV^t,$$

de lo cual se obtiene que

$$SV^tUS = S \qquad \text{o} \qquad SV^tU = I_r.$$

Tomando $R^* = SV^t$ y $Q = U$, se llega a

$$A = QR^* \qquad \text{con} \qquad R^*Q = I_r.$$

■

Teorema 8.5 *Toda matriz cuadrada real A de tamaño $n \times n$ que pueda expresarse en la forma:*

$$A = QQ^t, \tag{8.1}$$

donde Q es una matriz real de tamaño $n \times m$ $(m < n)$ con columnas ortonormales en $\mathbb{R}^m$, satisface lo siguiente:

(1) *A es simétrica e idempotente.*

(2) $A\big(I_n - A\big) = \big(I_n - A\big)A = \mathrm{O}\,.$

(3) $\big(I_n - A\big)Q = \mathrm{O}\,.$

Demostración.

(1) Si $A = QQ^t$, entonces

$$\begin{aligned} A^t &= \big(QQ^t\big)^t = \big(Q^t\big)^t Q^t = QQ^t = A \\ A^2 &= \big(QQ^t\big)\big(QQ^t\big) = Q\big(Q^tQ\big)Q^t = QI_mQ^t = QQ^t = A. \end{aligned}$$

(2) $A\big(I_n - A\big) = A - A^2 = \mathrm{O}$, e igual para el otro caso.

(3) $\big(I_n - A\big)Q = Q - \big(QQ^t\big)Q = Q - QI_m = \mathrm{O}\,.$ ■

Ejemplo 8.2 *Encuentre una factorización QR de la matriz:*

$$A = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Solución.

Denotemos las columnas de A por

$$\vec{x}_1 = -\frac{1}{3}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \qquad \vec{x}_2 = -\frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \qquad \vec{x}_3 = -\frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

Apliquemos el algoritmo de Gram-Schmidt al conjunto $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$, el cual es una base para el espacio generado por las columnas de A. Como $\|\vec{x}_1\| = \frac{1}{3}\sqrt{6}$, se hace $\vec{v}_1 = \dfrac{\vec{x}_1}{\|\vec{x}_1\|} = -\dfrac{1}{\sqrt{6}}\begin{bmatrix}-2\\1\\1\end{bmatrix}$. Por otra parte,

$$\vec{v}_2' = \vec{x}_2 - (\vec{x}_2 \cdot \vec{v}_1)\vec{v}_1 = \frac{1}{3}\begin{bmatrix}-1\\2\\-1\end{bmatrix} + \frac{1}{2}\frac{1}{3}\begin{bmatrix}2\\-1\\-1\end{bmatrix} = \frac{3}{6}\begin{bmatrix}0\\1\\-1\end{bmatrix}.$$

Entonces, $\left\|\vec{v}_2'\right\| = \frac{1}{2}\sqrt{2}$ y $\vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\-1\end{bmatrix}$. Por último,

$$\vec{v}_3' = \vec{x}_3 - (\vec{x}_3 \cdot \vec{v}_1)\vec{v}_1 - (\vec{x}_3 \cdot \vec{v}_2)\vec{v}_2 = \begin{bmatrix}-\frac{1}{3}\\-\frac{1}{3}\\\frac{2}{3}\end{bmatrix} + \frac{1}{6}\begin{bmatrix}2\\-1\\-1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\1\\-1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix};$$

esto sucede porque $\vec{v}_3$ es una combinación lineal de $\vec{v}_1$ y $\vec{v}_2$. Se puede verificar que $\vec{v}_1$ y $\vec{v}_2$ forman una base ortonormal para el espacio generado por las columnas de A, ya que $\vec{v}_1 \cdot \vec{v}_2 = 0$. Entonces, formamos la matriz

$$Q = \begin{bmatrix}\vec{v}_1 & \vec{v}_2\end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix}2 & 0\\-1 & \sqrt{3}\\-1 & -\sqrt{3}\end{bmatrix}.$$

Luego, la matriz R es

$$\begin{aligned} R = Q^t A &= \frac{1}{\sqrt{6}}\begin{bmatrix}2 & -1 & -1\\0 & \sqrt{3} & -\sqrt{3}\end{bmatrix}\begin{bmatrix}\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3}\\-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3}\\-\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}\end{bmatrix}\\ &= \frac{1}{\sqrt{6}}\begin{bmatrix}2 & -1 & -1\\0 & \sqrt{3} & -\sqrt{3}\end{bmatrix}. \end{aligned}$$

Nótese que $R = Q^t$, por lo tanto, la matriz A es idempotente.

Definición 8.2 Matriz de Householder

Una matriz de Householder viene dada por

$$H\big(\vec{v}\big) = I_n - 2\vec{v}\vec{v}^t, \tag{8.2}$$

donde $\vec{v} \in \mathbb{R}^n$ *y* $\vec{v}^t\vec{v} = 1$.

Teorema 8.6 *Las matrices de Householder son simétricas y ortogonales.*

Demostración.

Veamos primero que $H(\vec{v})$ es simétrica

$$\left[H(\vec{v})\right]^t = \left[I_n - 2\vec{v}\vec{v}^t\right]^t = I_n - 2\vec{v}\vec{v}^t = H(\vec{v}).$$

Ahora, se mostrará que $\left[H(\vec{v})\right]^t H(\vec{v}) = I_n$

$$\begin{aligned}\left[H(\vec{v})\right]^t H(\vec{v}) &= \left[I_n - 2\vec{v}\vec{v}^t\right]^2 = I_n - 4\vec{v}\vec{v}^t + 4\vec{v}\vec{v}^t\vec{v}\vec{v}^t \\ &= I_n - 4\vec{v}\vec{v}^t + 4\vec{v}\vec{v}^t = I_n,\end{aligned}$$

y se tiene lo que se quería demostrar. ■

8.1.1 Factorización QR por reflexiones de Householder

Una secuencia de transformaciones de Householder puede utilizarse para calcular la "factorización QR" de una matriz $A = [a_{ij}]$ de tamaño $m \times n$, ya que es posible escoger la matriz de Householder de manera que el vector elegido quede con una única componente no nula tras ser transformado (es decir, premultiplicando por la matriz de Householder). La manera de construir la matriz de Householder es la siguiente:

Sean $\vec{x} \in \mathbb{R}^m$ y $\vec{e}_1$ el primer vector unitario de $\mathbb{R}^m$, entonces se define

$$\begin{aligned}\vec{u} &= \vec{x} - s\|\vec{x}\|\vec{e}_1, & s &= \operatorname{sgn}(x_1) \\ H(\vec{v}) &= I_m - 2\vec{v}\vec{v}^t & \vec{v} &= \frac{\vec{u}}{\|\vec{u}\|},\end{aligned} \tag{8.3}$$

donde $\operatorname{sgn}(\cdot)$ denota la función signo y x_1 es el primer elemento de $\vec{x}$. Nótese que $\vec{v}$ es un vector unitario, luego $H(\vec{v})$ es una matriz de Householder y satisface que

$$\begin{aligned}\vec{x}^t H(\vec{v}) &= \vec{x}^t - \vec{u}^t = s||\vec{x}||\vec{e}_1^t \\ H(\vec{v})(A)_{\cdot k} &= (A)_{\cdot k} - \beta_k\vec{u},\end{aligned}$$

donde $(A)_{\cdot k}$ denota la k-ésima columna de A y

$$\beta_k = \frac{\alpha_k}{||\vec{x}||} \qquad \text{con} \qquad \alpha_k = \frac{\vec{x}^t(A)_{\cdot k} - s||\vec{x}||a_{1k}}{||\vec{x}|| - |x_1|}.$$

Este algoritmo se puede emplear para transformar gradualmente los vectores columna de la matriz A en una matriz triangular superior. En primer

lugar, se premultiplica A por la matriz de Householder $Q_1 = H\big(\vec{v}\big)$, que se obtiene eligiendo $\vec{x} = \big(A\big)_{.1}$. Esto produce una matriz Q_1A con ceros en la primera columna (excepto el primer elemento de la fila).

$$\begin{aligned} Q_1A &= H\big(\vec{v}\big)\left[\big(A\big)_{.1} \quad \big(A\big)_{.2} \ldots \big(A\big)_{.n}\right] \\ &= \left[s||\vec{x}||\vec{e}_1 \quad \big(A\big)_{.2} - \beta_2\vec{u} \quad \ldots \quad \big(A\big)_{.n} - \beta_n\vec{u}\right] \\ &= \begin{bmatrix} s||\vec{x}|| & r_{12} & \ldots & r_{1n} \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{bmatrix}. \end{aligned}$$

El procedimiento se puede repetir para A_2 (la cual se forma eliminando la primera fila y columna de A, respectivamente), obteniéndose así una nueva matriz de Householder Q_2'. Puesto que Q_2' es de tamaño menor que Q_1, para lograr que esta matriz opere con Q_1A en lugar de A_2, se necesita aumentarla hacia arriba a la izquierda, colocando un uno en la diagonal, o en general

$$Q_k = \begin{bmatrix} I_{k-1} & 0 \\ 0 & Q_k' \end{bmatrix}.$$

Si se repite el proceso r veces, donde $r = \text{mín}\{m-1, n\}$, entonces

$$R = Q_r \cdots Q_2Q_1A,$$

es una matriz triangular superior. Luego, tomando

$$Q = Q_1Q_2 \cdots Q_r,$$

se llega a $A = QR$: la descomposición QR de la matriz A.

Este método tiene una estabilidad numérica mayor que la del método de Gram-Schmidt empleado anteriormente. Aunque cabe aclarar que encontrar la factorización QR de una matriz de tamaño $m \times n$ mediante este método es extenso si se hace a mano.

Ejemplo 8.3 *Encontrar la descomposición QR de la matriz dada en el Ejemplo 8.2 mediante el uso de matrices de Householder.*

Solución.

Para encontrar la primera matriz de Householder, se emplea la primera columna de la matriz A, es decir, $\big(A\big)_{.1} = \frac{1}{3}\,(2,-1,-1)^t$. Luego,

$$\vec{x} \;= \big(A\big)_{.1} = \frac{1}{3}\,(2,-1,-1)^t \qquad \text{y} \qquad ||\vec{x}|| \;= \frac{1}{3}\sqrt{6}.$$

Usando la expresión (8.3), se tiene que

$$\vec{u} \;= \frac{1}{3}\left(2-\sqrt{6},-1,-1\right)^t \quad \text{y} \quad \vec{v} \;= \frac{1}{2\sqrt{3-\sqrt{6}}}\left(2-\sqrt{6},-1,-1\right)^t,$$

entonces

$$Q_1 = H\big(\vec{v}\big) = I_3 - \frac{2}{4(3-\sqrt{6})}\begin{bmatrix} 2-\sqrt{6} \\ -1 \\ -1 \end{bmatrix}\begin{bmatrix} 2-\sqrt{6} & -1 & -1 \end{bmatrix}$$

$$= I_3 - \frac{1}{6}\begin{bmatrix} 6-2\sqrt{6} & \sqrt{6} & \sqrt{6} \\ \sqrt{6} & 3+\sqrt{6} & 3+\sqrt{6} \\ \sqrt{6} & 3+\sqrt{6} & 3+\sqrt{6} \end{bmatrix} = \frac{-1}{\sqrt{6}}\begin{bmatrix} -2 & 1 & 1 \\ 1 & 1-\frac{\sqrt{6}}{2} & 1+\frac{\sqrt{6}}{2} \\ 1 & 1+\frac{\sqrt{6}}{2} & 1-\frac{\sqrt{6}}{2} \end{bmatrix}.$$

Ahora se calcula

$$R = [r_{ij}] = Q_1^t A = \frac{1}{6}\begin{bmatrix} 2\sqrt{6} & -\sqrt{6} & -\sqrt{6} \\ 0 & 3 & -3 \\ 0 & -3 & 3 \end{bmatrix},$$

con lo que ya casi se tiene una matriz triangular. Solo se necesita hacer cero el elemento r_{32}, tomando la submatriz principal bajo el elemento r_{11} y aplicando de nuevo el proceso a

$$A_2 = M_{11} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Mediante el mismo método que antes, se obtiene la matriz de Householder:

$$Q_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & -1 \end{bmatrix}.$$

Finalmente, se obtiene

$$Q \;= \frac{1}{\sqrt{6}}\begin{bmatrix} 2 & 0 & \sqrt{2} \\ -1 & -\sqrt{3} & \sqrt{2} \\ -1 & \sqrt{3} & \sqrt{2} \end{bmatrix} \qquad \text{y} \qquad R \;= \frac{1}{\sqrt{6}}\begin{bmatrix} 2 & -1 & -1 \\ 0 & -\sqrt{3} & \sqrt{3} \\ 0 & 0 & 0 \end{bmatrix}.$$

La matriz $Q = Q_1Q_2$ es ortogonal y $R = Q^tA$ es triangular superior, de forma que $A = QR$ es la descomposición QR buscada.

Teorema 8.7 *Si A es simétrica e idempotente y P es ortogonal, entonces P^tAP es idempotente.*

Demostración.
Si P es ortogonal, entonces

$$(P^tAP)(P^tAP) = P^tA(PP^t)AP = P^t(AA)P = P^tAP. \quad \blacksquare$$

Teorema 8.8 *Sea A una matriz simétrica e idempotente de tamaño $n \times n$ con rango r, entonces existe una matriz ortogonal Q tal que $Q^tAQ = D_r$, donde D_r es una matriz diagonal con r elementos iguales a uno y los elementos restantes $n - r$ de la diagonal iguales a cero.*

Demostración.
Este se sigue inmediatamente del Teorema 2.29. $\blacksquare$

Teorema 8.9 *Si A es una matriz idempotente de tamaño $n \times n$, entonces su forma de Jordan $\mathcal{J} = P^{-1}AP$ satisface que $\mathcal{J}^2 = \mathcal{J}$.*

Demostración.
Queda como ejercicio para el lector. $\blacksquare$

Teorema 8.10 *Si $A = [a_{ij}]$ es una matriz simétrica e idempotente y si el i-ésimo elemento de la diagonal es cero, entonces los elementos de la i-ésima fila y la i-ésima columna son todos idénticamente cero.*

Demostración.
Puesto que $A = A^2$, se tiene que el i-ésimo elemento de la diagonal de A es

$$a_{ii} = \sum_{j=1}^{n} a_{ij}a_{ji}.$$

Pero como A es simétrica $a_{ij} = a_{ji}$,

$$a_{ii} = \sum_{j=1}^{n} a_{ij}^2.$$

Luego, si $a_{ii} = 0$, entonces $a_{ij} = 0$ (para $j = 1, 2, \ldots, n$), es decir, los elementos de la i-ésima fila son todos cero. Y como $A = A^t$ se tiene que los elementos de la i-ésima columna son también todos cero. ■

Teorema 8.11 *El producto de dos matrices simétricas e idempotentes es idempotente si el producto de las dos matrices es conmutativo.*

Demostración.

Si $AB = BA$, entonces

$$(AB)(AB) = (AB)(BA) = A(BA) = A(AB) = AB. \blacksquare$$

Teorema 8.12 *Sea A una matriz real de tamaño $m \times n$ $(m > n)$ con $\rho\big(A\big) = n$, entonces la matriz $C = A\big(A^tA\big)^{-1}A^t$ es simétrica e idempotente.*

Demostración.

La matriz $C = A\big(A^tA\big)^{-1}A^t$ es simétrica, ya que

$$C^t = \big[A\big(A^tA\big)^{-1}A^t\big]^t = \big(A^t\big)^t\big[\big(A^tA\big)^{-1}\big]^tA^t = A\big[\big(A^tA\big)^t\big]^{-1}A^t = C.$$

Además, es idempotente, pues

$$C^2 = \big[A\big(A^tA\big)^{-1}A^t\big]\big[A\big(A^tA\big)^{-1}A^t\big] = AI_n\big(A^tA\big)^{-1}A^t = C.$$

Nótese que la matriz A^tA es no singular, pues A es de rango completo columna y $\rho\big(A^tA\big) = \rho\big(A\big)$. ■

Teorema 8.13 *Sea A una matriz simétrica e idempotente de tamaño $n \times n$, entonces*

$$\rho(A) = \mathrm{tr}(A).$$

Demostración.

Por el Teorema 8.8, existe una matriz ortogonal Q tal que $A = QD_rQ^t$. Luego, se tiene que

$$\mathrm{tr}(A) = \mathrm{tr}(QD_rQ^t) = \mathrm{tr}(D_rQ^tQ) = \mathrm{tr}(D_r) = r = \rho(A). \blacksquare$$

Teorema 8.14 *Todas las matrices simétricas idempotentes de rango incompleto son semidefinidas positivas.*

Demostración.
Queda como ejercicio para el lector. ■

Teorema 8.15 *Sea A una matriz real de tamaño $n \times n$, entonces A es simétrica e idempotente si y solo si*

$$\rho(A) + \rho(I_n - A) = n.$$

Demostración.
Supongamos que

$$\rho(A) + \rho(I_n - A) = n,$$

y sea $\mathcal{R}(A)$ el espacio de los renglones de A. Veamos que

$$\mathbb{R}^n = \mathcal{R}(A) \oplus \mathcal{R}(I_n - A).$$

Obsérvese que

$$\begin{aligned} n &= \dim\big(\mathbb{R}^n\big) = \dim\big\{\mathcal{R}(A) \oplus \mathcal{R}(I_n - A)\big\} \\ &= \dim\big\{\mathcal{R}(A)\big\} + \dim\big\{\mathcal{R}(I_n - A)\big\} - \dim\big\{\mathcal{R}(A) \cap \mathcal{R}(I_n - A)\big\} \\ &= \rho(A) + \rho(I_n - A) - \dim\big\{\mathcal{R}(A) \cap \mathcal{R}(I_n - A)\big\} \\ &= n - \dim\big\{\mathcal{R}(A) \cap \mathcal{R}(I_n - A)\big\}. \end{aligned}$$

Esto implica que $\dim\big\{\mathcal{R}(A) \cap \mathcal{R}(I_n - A)\big\} = 0$, de lo cual se tiene que

$$\mathcal{R}(A) \cap \mathcal{R}(I_n - A) = \vec{0}.$$

Por consiguiente, $\mathbb{R}^n = \mathcal{R}(A) \oplus \mathcal{R}(I_n - A)$, y esto exige que

$$A\big(I_n - A\big) = \mathrm{O}\,.$$

Supongamos que no es así, entonces existen vectores no nulos $\vec{u}$ y $\vec{v}$ en $\mathbb{R}^n$ tales que

$$A\big(I_n - A\big)\vec{u} = \vec{v}.$$

Luego, $\vec{v} \in \mathcal{R}(A)$, pero como $A\big(I_n - A\big) = \big(I_n - A\big)A$, se tiene que

$$\big(I_n - A\big)A\vec{u} = \vec{v}.$$

Esto implica que $\vec{v} \in \mathcal{R}(I_n - A)$ y se llega a una contradicción. Por lo tanto, $A\big(I_n - A\big) = \mathrm{O}$ o $A^2 = A$. Esto completa la prueba. ■

Ejemplo 8.4 *Determine el rango de la matriz asociada a la forma cuadrática del Ejemplo 5.5.*

Solución.

La matriz asociada a la forma cuadrática dada en el Ejemplo 5.5 era $I_n - \overline{J}_n$. Veamos si es simétrica e idempotente.

$$\begin{aligned}\big(I_n - \overline{J}_n\big)^t &= \left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)^t = \left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\\ \big(I_n - \overline{J}_n\big)^2 &= \left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t\right)\\ &= I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^t + \frac{1}{n^2}\mathbf{1}\underbrace{\mathbf{1}^t\mathbf{1}}\mathbf{1}^t = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t = I_n - \overline{J}_n.\end{aligned}$$

Luego, por el teorema anterior se tiene que

$$\rho\big(I_n - \overline{J}_n\big) = n - \rho\big(\overline{J}_n\big) = n - 1,$$

pues la matriz $\overline{J}_n$ tiene únicamente una fila linealmente independiente.

Teorema 8.16 *Sean A_1 y A_2 dos matrices cuadradas del mismo tamaño y $A = A_1 + A_2$, entonces las siguientes condiciones son equivalentes:*

(1) A es simétrica e idempotente y $\rho(A) = \rho(A_1) + \rho(A_2)$.

(2) A_1 y A_2 son simétricas e idempotentes y $A_1A_2 = A_2A_1 = \mathrm{O}$.

Demostración.

Supongamos que (2) es verdadero, entonces

$$\begin{aligned}A^2 &= \big(A_1 + A_2\big)\big(A_1 + A_2\big)\\ &= A_1^2 + A_2^2 + A_1A_2 + A_2A_1 = A_1 + A_2.\end{aligned}$$

Puesto que A, A_1 y A_2 son idempotentes,

$$\begin{aligned}\rho(A) &= \mathrm{tr}(A) = \mathrm{tr}\big(A_1 + A_2\big) = \mathrm{tr}\big(A_1\big) + \mathrm{tr}\big(A_2\big)\\ &= \rho\big(A_1\big) + \rho\big(A_2\big).\end{aligned}$$

Ahora, supongamos que (1) es verdadero, por el Teorema 8.15,

$$\begin{aligned} n &= \rho(A) + \rho\big(I_n - A\big) = \rho\big(A_1\big) + \rho\big(A_2\big) + \rho\big(I_n - A\big) \\ &\geq \rho\big(A_1\big) + \rho\big[A_2 + \big(I_n - A\big)\big] = \rho\big(A_1\big) + \rho\big(I_n - A_1\big) \\ &\geq \rho\big[A_1 + \big(I_n - A_1\big)\big] = \rho\big(I_n\big) = n. \end{aligned}$$

Por consiguiente, $\rho\big(A_1\big)+\rho\big(I_n-A_1\big) = n$ y de nuevo por el Teorema 8.15, se tiene que A_1 es idempotente; de manera análoga, se puede mostrar que A_2 es idempotente. Ahora demostremos que $A_1A_2 = A_2A_1 = \mathrm{O}$. Dado que A, A_1 y A_2 son idempotentes y $A = A_1 + A_2$, multiplicando ambos lados por A, se obtiene que

$$\begin{aligned} A &= A^2 = \big(A_1 + A_2\big)\big(A_1 + A_2\big) \\ &= A_1^2 + A_2^2 + A_1A_2 + A_2A_1 = \big(A_1 + A_2\big) + A_1A_2 + A_2A_1 \\ &= A + A_1A_2 + A_2A_1. \end{aligned}$$

Esto implica que

$$A_1A_2 + A_2A_1 = \mathrm{O}, \qquad \text{es decir,} \qquad A_1A_2 = -A_2A_1.$$

Por otra parte, el hecho de que $\rho(A) = \rho(A_1) + \rho(A_2)$ implica que

$$\mathcal{R}(A_1) \cap \mathcal{R}(A_2) = \{\vec{0}\}.$$

Este hecho unido con $A_1A_2 = -A_2A_1$ da $A_1A_2 = \mathrm{O}$. ■

Corolario 8.16.1 *Sean A_1, A_2 dos matrices de tamaño $n \times n$ tal que:*

$$A_1 + A_2 = I_n,$$

entonces las condiciones dadas en el Teorema 8.16 se cumplen.

Demostración.

Queda como ejercicio para el lector. ■

Una generalización del Teorema 8.16 en el que se involucran más de dos matrices se presenta a continuación.

Teorema 8.17 Teorema de Cochran

Sean $A_1, A_2, \ldots, A_m$ una colección de m matrices de tamaño $n \times n$ y $A = \sum_{i=1}^{m} A_i$. Considere las siguientes condiciones:

(1) Cada A_i es simétrica, idempotente y conmuta con A.

(2) $A_iA_j = \mathrm{O}$ para toda $i \neq j$ y $\rho\big(A_i^2\big) = \rho\big(A_i\big)$ para toda i.

(3) A es simétrica e idempotente.

(4) $\rho(A) = \sum_{i=1}^{m} \rho\big(A_i\big)$.

Entonces, cualquiera dos de las condiciones (1), (2) *y* (3) *implican la validez de la condición* (4). *Además, las condiciones* (3) *y* (4) *implican la validez del resto de las condiciones.*

Demostración.

Suponga que (1) y (2) son dadas. Como $A = \sum_{i=1}^{m} A_i$, es claro que es idempotente. Puesto que A y $A_1, A_2, \ldots, A_m$ son todas idempotentes,

$$\rho(A) = \mathrm{tr}(A) = \sum_{i=1}^{m} \mathrm{tr}\big(A_i\big) = \sum_{i=1}^{m} \rho\big(A_i\big).$$

Así, la condición (4) es verdadera.

Suponga que (2) y (3) son dadas. El cómputo de A^2 produce

$$A^2 = \sum_{i=1}^{m} A_i^2, \qquad \text{para} \qquad 1 \leq i \leq m.$$

Nótese que

$$AA_i = A_iA = A_i^2 \qquad \text{y} \qquad A^2A_i = A_iA^2 = A_i^3,$$

como A es idempotente, se tiene que $A_i^2 = A_i^3$, lo cual implica que $A_i^2\big(I_n - A_i\big) = \mathrm{O}$. La condición $\rho\big(A_i\big) = \rho\big(A_i^2\big)$ es equivalente a la siguiente afirmación:

$$\dim\big\{\mathcal{R}\big(A_i\big)\big\} = \dim\big\{\mathcal{R}\big(A_i^2\big)\big\}.$$

Puesto que $\mathcal{R}\big(A_i^2\big) \subset \mathcal{R}\big(A_i\big)$, se tiene que $\mathcal{R}\big(A_i\big) = \mathcal{R}\big(A_i^2\big)$. Por consiguiente, existe una matriz D no singular tal que $A_i = DA_i^2$. Por lo tanto, $A_i^2\big(I_n - A\big) = \mathrm{O}$ implica que $A_i\big(I_n - A\big) = \mathrm{O}$ de lo cual se concluye que A_i es idempotente. Así, la condición (1) es verdadera y se sigue la (4).

Supongamos que (3) y (4) son válidas. Para $i \neq j$, sea $B = A_i + A_j$ y $C = A - B$, por (4)

$$\begin{aligned}\sum_{i=1}^{m} \rho\big(A_i\big) \ &= \rho(A) = \rho\big(B + C\big) \\ &\leq \rho(B) + \rho(C) \leq \sum_{i=1}^{m} \rho\big(A_i\big).\end{aligned}$$

De esto, se tiene que $\rho(A) = \rho(B) + \rho(C)$, por otra parte,

$$\begin{aligned} n \ &= \rho\big(I_n\big) = \rho\big(B + I_n - B\big) \leq \rho(B) + \rho\big(I_n - B\big) \\ &= \rho(B) + \rho\big(I_n - A + C\big) \leq \rho(B) + \rho\big(I_n - A\big) + \rho(C) \\ &= \rho(A) + \rho\big(I_n - A\big) = n.\end{aligned}$$

Por lo tanto, $\rho(B) + \rho\big(I_n - B\big) = n$ y por el Teorema 8.15, B es idempotente. Así se tiene que $A_i + A_j$ es idempotente y $\rho(B) = \rho\big(A_i\big) + \rho\big(A_j\big)$. Por el Teorema 8.16, $A_i A_j = \mathrm{O}$ y A_i y A_j son idempotentes. Así, (2) y (3) se obtienen de una vez.

Suponga que (1) y (2) se cumplen. Es obvio que (4) se sigue aprovechando la conexión entre rango y traza para matrices idempotentes. Por lo tanto, se tiene que (4) es válido y (3) se sigue ahora de lo que se ha establecido anteriormente. Esto completa la prueba. ■

Corolario 8.17.1 *Sean $A_1, A_2, \ldots, A_m$ una colección de matrices simétricas e idempotentes de tamaño $n \times n$ tal que*

$$\sum_{i=1}^{m} A_i = I_n,$$

entonces las condiciones dadas en el Teorema 8.17 se cumplen. En este caso, las condiciones (1) *y* (2) *son equivalentes.*

Demostración.

Queda como ejercicio para el lector. ■

Teorema 8.18 *Sean $A_1, A_2, \ldots, A_m$ una colección de matrices simétricas e idempotentes de tamaño $n \times n$. Una condición necesaria y suficiente para que exista una matriz P ortogonal tal que $P^t A_1 P, P^t A_2 P, \ldots,$ $P^t A_m P$ sean todas diagonales es que $A_i A_j = A_j A_i$ para toda i y j.*

Demostración.
Queda como ejercicio para el lector. ■

Ejercicios 8.1

1. Obtenga condiciones para los elementos de las matrices idempotentes de tamaño 2×2. ¿Se puede generalizar a cualquier dimensión?

2. Muestre que si A es idempotente, entonces A^t es idempotente.

3. Sea X una matriz de tamaño $m \times n$ $(m > n)$ y rango n. Demuestre que la matriz $H = X\big(X^tX\big)^{-1}X^t$ es una matriz simétrica e idempotente. Obtenga la inversa de $I_m - H$.

4. Suponga que $KA = \mathrm{O}$ con K idempotente. Defina $G = \big(A - K\big)^{-1}$. Pruebe que:

$$(i)\ AG = I - K, \quad (ii)\ AGA = A \quad \text{y} \quad (iii) \quad AGK = \mathrm{O}\,.$$

8.2 Productos especiales

En esta sección se presentan nuevos conceptos de operaciones entre matrices, y entre otros, algunos productos especiales entre matrices.

Definición 8.3 Producto Kronecker

Dadas $A = [a_{ij}]$ *y* $B = [b_{kl}]$ *matrices reales de tamaño* $m \times n$ *y* $p \times q$, *respectivamente, el producto Kronecker entre ellas, operación que se denotará por* $A \otimes B$, *es una matriz* $C = [c_{ij}] = [a_{ij}B]_{ij}$ *de tamaño* $mp \times nq$, *dada por*

$$A \otimes B = [a_{ij}B]_{ij} = \begin{bmatrix} a_{11}\,B & \dots & a_{1j}\,B & \dots & a_{1n}\,B \\ \vdots & \ddots & \vdots & \dots & \vdots \\ a_{i1}\,B & \dots & a_{ij}\,B & \dots & a_{in}\,B \\ \vdots & \ddots & \vdots & \dots & \vdots \\ a_{m1}\,B & \dots & a_{mj}\,B & \dots & a_{mn}\,B \end{bmatrix}, \tag{8.4}$$

donde cada submatriz $a_{ij}B$ es de tamaño $p \times q$. Este producto tiene sentido tanto para matrices como para vectores.

Ejemplo 8.5 *Considere las matrices dadas en el Ejemplo 2.8, determine $A \otimes B$ y $B \otimes A$.*

Solución.

Primero, se realiza $A \otimes B$:

$$A \otimes B = \begin{bmatrix} 1\,B & 4\,B \\ 4\,B & 1\,B \end{bmatrix} = \begin{bmatrix} 1\begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix} & 4\begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix} \\ 4\begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix} & 1\begin{bmatrix} 1 & 1 \\ 1 & -14 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 4 & 4 \\ 1 & -14 & 4 & -56 \\ 4 & 4 & 1 & 1 \\ 4 & -56 & 1 & -14 \end{bmatrix}.$$

Ahora, se realiza $B \otimes A$:

$$B \otimes A = \begin{bmatrix} 1\,A & 1\,A \\ 1\,A & -14\,A \end{bmatrix} = \begin{bmatrix} 1\begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} & 1\begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} \\ 1\begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} & -14\begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1 & 4 & 1 & 4 \\ 4 & 1 & 4 & 1 \\ 1 & 4 & -14 & -56 \\ 4 & 1 & -56 & -14 \end{bmatrix},$$

nótese que $A \otimes B \neq B \otimes A$.

Teorema 8.19 Propiedades del producto Kronecker

Sean A, B y C matrices cualesquiera y $\vec{u}$ y $\vec{v}$ dos vectores columna. Entonces:

1. *$\alpha \otimes A = A \otimes \alpha = \alpha A$, para cualquier escalar α.*

2. $\big(\alpha A\big) \otimes \big(\beta B\big) = \alpha\beta\big(A \otimes B\big)$, *para cualesquiera escalares* α, β.

3. $A \otimes B \neq B \otimes A$

4. $\big(A \otimes B\big) \otimes C = A \otimes \big(B \otimes C\big)$

5. $\big(A + B\big) \otimes C = A \otimes C + B \otimes C$, *si* A *y* B *son del mismo tamaño.*

6. $A \otimes \big(B + C\big) = A \otimes B + A \otimes C$, *si* B *y* C *son del mismo tamaño.*

7. $\big(A \otimes B\big)^t = A^t \otimes B^t$.

8. $\rho\big(A \otimes B\big) = \rho\big(A\big)\rho\big(B\big)$.

9. $\mathrm{tr}\big(A \otimes B\big) = \mathrm{tr}\big(A\big)\,\mathrm{tr}\big(B\big)$, *si* A *y* B *son cuadradas.*

10. *Sean* A, B, C *y* D *matrices de tamaño* $m \times h$, $p \times k$, $h \times n$ *y* $k \times q$, *respectivamente, entonces:*

$$\big(A \otimes B\big)\big(C \otimes D\big) = (AC) \otimes (BD). \tag{8.5}$$

11. $\big(A \otimes B\big)^{-1} = A^{-1} \otimes B^{-1}$, *si* A *y* B *son matrices cuadradas invertibles de tamaño* $m \times m$ *y* $n \times n$, *respectivamente, y* $\big(A \otimes B\big)$ *es no singular.*

12. *Si* A *y* B *son matrices simétricas de tamaño* $m \times m$ *y* $n \times n$, *respectivamente, entonces* $A \otimes B$ *es simétrica.*

13. *Si* A *y* B *son matrices de tamaño* $m \times n$ *y* $p \times q$, *respectivamente, con descomposiciones en valores singulares* $U_A S_A V_A^t$ *y* $U_B S_B V_B^t$, *entonces:*

$$\big(U_A \otimes U_B\big)\big(S_A \otimes S_B\big)\big(V_A^t \otimes V_B^t\big)$$

produce una descomposición en valores singulares de $A \otimes B$ (después de un simple reordenamiento de los elementos de la diagonal de $S_A \otimes S_B$ y los correspondientes vectores singulares).

14. *Si A y B son matrices simétricas de tamaño $n \times n$, entonces usando la descomposición espectral de cada matriz, el producto Kronecker se puede expresar como sigue:*

$$A \otimes B = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mu_j \big(\vec{u}_i \vec{u}_i^t \otimes \vec{v}_j \vec{v}_j^t\big),$$

donde λ_i, $\mu_j \in \mathbb{R}$ son los valores propios y $\vec{u}_i$, $\vec{v}_j$ son los vectores propios normalizados de las matrices A y B, respectivamente.

15. $\vec{u}\vec{v}^t = \vec{u} \otimes \vec{v}^t = \vec{v}^t \otimes \vec{u}$.

Demostración.

1. $-$ 3. Quedan como ejercicio para el lector.

4. Por la Definición 8.3, se tiene que

$$A \otimes B \; = \; [a_{ij} B]_{ij} \, ,$$

luego,

$$\big(A \otimes B\big) \otimes C \; = \; \Big[\big(a_{ij} B\big)_{ij}\Big] \otimes C = a_{ij} \big(B \otimes C\big)_{ij},$$

por lo tanto, $\big(A \otimes B\big) \otimes C = A \otimes \big(B \otimes C\big)$.

5. Sea $E = A + B = [e_{ij}]$. Por la Definición 8.3, se tiene que

$$\begin{aligned} E \otimes C \; &= \; [e_{ij} C]_{ij} = \big[(a_{ij} + b_{ij}) C\big]_{ij} \\ &= \; [a_{ij} C]_{ij} + [b_{ij} C]_{ij} = A \otimes C + B \otimes C. \end{aligned}$$

6. $-$ 8. Quedan como ejercicio para el lector.

9. Usando (8.4), cuando $m = n$, se tiene que

$$\begin{aligned}\operatorname{tr}\big(A \otimes B\big) &= \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n} a_{ii} \operatorname{tr}\big(B\big) = \Big[\sum_{i=1}^{n} a_{ii}\Big] \operatorname{tr}\big(B\big) \\ &= \operatorname{tr}\big(A\big) \operatorname{tr}\big(B\big).\end{aligned}$$

10. Usando (8.4), cuando $m = n$, se tiene que

$$\begin{aligned}\big(A \otimes B\big)\big(C \otimes D\big) &= \sum_{l=1}^{h} [a_{il}B]_{il} \, [c_{lj}D]_{lj} = \Big[\sum_{l=1}^{h} a_{il}c_{lj}\Big] BD \\ &= (A)_{i\cdot}(C)_{\cdot j}\big(BD\big) = \Big[\big(AC\big)_{ij}\big(BD\big)\Big]_{ij} \\ &= \big(AC\big) \otimes \big(BD\big).\end{aligned}$$

11. Usando la expresión (8.5), note que

$$\big(A \otimes B\big)\big(A^{-1} \otimes B^{-1}\big) = I_m \otimes I_n = I_{mn}.$$

12. $-$ 15. Quedan como ejercicio para el lector. ■

Teorema 8.20 *Sea A una matriz de tamaño $m \times m$, con valores propios $\lambda_1, \lambda_2, \ldots, \lambda_m$ y B una matriz de tamaño $n \times n$, con valores propios $\mu_1, \mu_2, \ldots, \mu_n$, entonces el conjunto de mn valores propios de $A \otimes B$ es dado por $\big\{\lambda_i\mu_j : i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\big\}$.*

Demostración.

Por el Teorema 3.19, existen matrices no singulares P y Q tales que

$$P^{-1}AP = \mathcal{J}_A \qquad\qquad Q^{-1}BQ = \mathcal{J}_B,$$

donde $\mathcal{J}_A$ y $\mathcal{J}_B$ son matrices triangulares superiores con los valores propios de A y B como elementos de la diagonal, respectivamente. Los valores propios de $A \otimes B$ son los mismos de

$$\begin{aligned}\big(P \otimes Q\big)^{-1}\big(A \otimes B\big)\big(P \otimes Q\big) &= \big(P^{-1} \otimes Q^{-1}\big)\big(A \otimes B\big)\big(P \otimes Q\big) \\ &= \big(P^{-1}AP\big) \otimes \big(Q^{-1}BQ\big) = \mathcal{J}_A \otimes \mathcal{J}_B.\end{aligned}$$

Esta matriz es triangular superior ya que $\mathcal{J}_A$ y $\mathcal{J}_B$ son triangulares y, por lo tanto, los elementos diagonales de $\mathcal{J}_A \otimes \mathcal{J}_B$ son sus valores propios, los cuales vienen dados por

$$\big\{\lambda_i\mu_j : i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\big\}. \quad ■$$

Corolario 8.20.1 *Si A y B son matrices cuadradas de tamaño $m \times m$ y $n \times n$, respectivamente, entonces*

$$\det\big[A \otimes B\big] = \big[\det\big(A\big)\big]^n \big[\det\big(B\big)\big]^m. \tag{8.6}$$

Demostración.

Por el Teorema 3.19, existen matrices no singulares P y Q tales que

$$P^{-1}AP = \mathcal{J}_A \qquad\qquad Q^{-1}BQ = \mathcal{J}_B,$$

entonces los valores propios de $A \otimes B$ son los mismos de $\mathcal{J}_A \otimes \mathcal{J}_B$, luego

$$\begin{aligned}
\det\big[\mathcal{J}_A \otimes \mathcal{J}_B\big] &= \prod_{j=1}^{n}\prod_{i=1}^{m}\big(\lambda_i\mu_j\big) = \prod_{j=1}^{n}\mu_j^m\left(\prod_{i=1}^{m}\lambda_i\right) \\
&= \prod_{j=1}^{n}\mu_j^m \det\big(A\big) = \big(\det A\big)^n\left[\prod_{j=1}^{n}\mu_j\right]^m \\
&= \big(\det A\big)^n\big(\det B\big)^m. \ \blacksquare
\end{aligned}$$

Teorema 8.21 *Si A y B son matrices semidefinidas positivas de tamaño $m \times m$ y $n \times n$, respectivamente, entonces $A \otimes B$ también lo es.*

Demostración.

Queda como ejercicio para el lector. ■

Ahora, se considera el operador que transforma una matriz en un vector, el cual es conocido como el *operador Vec*.

Definición 8.4 Operador Vec

Sea $A = [a_{ij}]$ una matriz real de tamaño $m \times n$, si $(A)_{\cdot j}$ denota la j-ésima columna de A, entonces $\mathrm{vec}(A)$ *es el vector de tamaño $mn \times 1$ dado por*

$$\mathrm{vec}(A) = \begin{bmatrix} (A)_{\cdot 1} \\ (A)_{\cdot 2} \\ \vdots \\ (A)_{\cdot n} \end{bmatrix} \qquad \textit{con} \qquad (A)_{\cdot j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}. \tag{8.7}$$

Ejemplo 8.6 *Considere las matrices dadas en el Ejemplo 2.8, determine* $\text{vec}(A)$ *y* $\text{vec}(B)$.

Solución.

En este caso, $\text{vec}(A)$ y $\text{vec}(B)$ vienen dados por

$$\text{vec}(A) = \begin{bmatrix} 1 \\ 4 \\ 4 \\ 1 \end{bmatrix} \qquad \text{vec}(B) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -14 \end{bmatrix}.$$

Teorema 8.22 Propiedades del operador Vec

Sean A, B *y* C *matrices cualesquiera y* $\vec{u}$ *y* $\vec{v}$ *dos vectores columna. Entonces:*

1. $\text{vec}\big(\alpha A + \beta B\big) = \alpha\ \text{vec}(A) + \beta\ \text{vec}(B)$, *si* A *y* B *son matrices del mismo tamaño y para cualesquiera escalares* α, β.

2. $\text{tr}\big(A^t B\big) = \big[\text{vec}(A)\big]^t \text{vec}(B)$, *si* A *y* B *son matrices de tamaño* $m \times n$.

3. *Si* A, B *y* C *son matrices de tamaño* $m \times n$, $n \times p$ *y* $p \times q$, *respectivamente, entonces* $\text{vec}\big(ABC\big) = \big(C^t \otimes A\big)\,\text{vec}(B)$.

4. $\text{vec}\big(\vec{u}\big) = \text{vec}\big(\vec{u}^t\big) = \vec{u}$.

5. $\text{vec}\big(\vec{u}\,\vec{v}^t\big) = \vec{v} \otimes \vec{u}$.

Demostración.

Véase en Graybill (1983). ■

Definición 8.5 Producto Hadamard

Dadas $A = [a_{ij}]$ *y* $B = [b_{ij}]$, *matrices reales de tamaño* $m \times n$, *entonces el producto Hadamard entre ellas, operación que se denotará*

por $A \odot B$, es una matriz C de tamaño $m \times n$ cuyo elemento genérico c_{ij}, viene dado por:

$$A \odot B = C = [c_{ij}] = [a_{ij}b_{ij}], \qquad i = 1, 2, \ldots, m; j = 1, 2, \ldots, n. \quad (8.8)$$

Este producto tiene sentido tanto para matrices como para vectores.

Ejemplo 8.7 *Considere las matrices dadas en el Ejemplo 2.8, determine $A \odot B$.*

Solución.

Empleando la expresión (8.8), se resuelve $A \odot B$

$$A \odot B = \begin{bmatrix} a_{11}\, b_{11} & a_{12}\, b_{12} \\ a_{21}\, b_{21} & a_{22}\, b_{22} \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 4 & -14 \end{bmatrix}.$$

Teorema 8.23 Propiedades del producto Hadamard

Sean A, B y C matrices de tamaño $m \times n$, entonces:

1. $A \odot B = B \odot A$.
2. $\big(A \odot B\big) \odot C = A \odot \big(B \odot C\big)$.
3. $A \odot \big(B + C\big) = A \odot B + A \odot C$.
4. $\big(A + B\big) \odot C = A \odot C + B \odot C$.
5. $\big(A \odot B\big)^t = A^t \odot B^t$.
6. $\rho\big(A \odot B\big) \leq \rho\big(A\big)\rho\big(B\big)$.
7. *Si* $J_{mn} = \mathbf{1}_m \mathbf{1}_n^t$, *entonces* $A \odot J_{mn} = J_{mn} \odot A = A$.

8. *Si $m = n$ y D es una matriz diagonal de tamaño $n \times n$, entonces:*

$$\begin{aligned} &a) \quad (A \odot B)D = (AD) \odot B = A \odot (BD). \\ &b) \quad D(A \odot B) = (DA) \odot B = A \odot (DB). \end{aligned} \tag{8.9}$$

9. $\operatorname{vec}\big(A \odot B\big) = \big(\operatorname{vec} A\big) \odot \big(\operatorname{vec} B\big)$.

10. $\vec{u}\vec{v}^t \odot \vec{w}\vec{x}^t = \big(\vec{u} \odot \vec{w}\big)\big(\vec{v} \odot \vec{x}\big)^t$, *donde* $\vec{u}, \vec{w} \in \mathbb{R}^m$ *y* $\vec{v}, \vec{x} \in \mathbb{R}^n$.

11. *Si A y B son matrices simétricas de tamaño $n \times n$, entonces usando la descomposición espectral de cada matriz, el producto Hadamard se puede expresar como sigue:*

$$A \odot B = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mu_j \big(\vec{u}_i \odot \vec{v}_j\big)\big(\vec{u}_i \odot \vec{v}_j\big)^t,$$

donde λ_i, $\mu_j \in \mathbb{R}$ son los valores propios y $\vec{u}_i$, $\vec{v}_i$ son los vectores propios normalizados de las matrices A y B, respectivamente.

12. *Si A es una matriz simétrica de tamaño 2×2, el producto Hadamard de A por ella misma k-veces, es decir $A \odot A \odot \ldots \odot A = \bigodot_{j=1}^{k} A$, es*

$$\bigodot_{j=1}^{k} A = \Big(\frac{\beta}{2}\Big)^k \begin{bmatrix} \Big(\frac{\alpha}{\beta} + \cos 2\theta\Big)^k & \big(\operatorname{sen} 2\theta\big)^k \\ \big(\operatorname{sen} 2\theta\big)^k & \Big(\frac{\alpha}{\beta} - \cos 2\theta\Big)^k \end{bmatrix}, \tag{8.10}$$

donde $\alpha = \lambda_1 + \lambda_2$, $\beta = \lambda_1 - \lambda_2$, $\big(\lambda_1 > \lambda_2\big)$, con λ_i los valores propios de A y θ el ángulo de rotación de la forma cuadrática representada por A.

Demostración.

Puesto que algunas de las propiedades que se presentan en este material aparecen demostradas en algunos textos de álgebra lineal, solo se demostrarán las propiedades (8.9) y (8.10).

1. $-$ 7. Quedan como ejercicio para el lector.

8. Se verifica solo la parte a). Para probar la otra parte, se procede de manera análoga. Por la Definición 8.5, se tiene que

$$A \odot B = C \ = [c_{ij}] = [a_{ij}b_{ij}]\,, \quad i \ = 1, 2, \ldots, m; j = 1, 2, \ldots, n,$$

sea $E = CD = [e_{ij}]$, donde

$$e_{ij} \ = \sum_{k=1}^{n} c_{ik}d_{kj},$$

pero como D es una matriz diagonal, entonces $d_{kj} = 0$ si $k \neq j$, luego $c_{ik}d_{kj} = 0$ cuando $k \neq j$. Por lo tanto,

$$e_{ij} \ = c_{ij}d_{jj},$$

y al sustituir c_{ij}, se obtiene que

$$e_{ij} \ = a_{ij}b_{ij}d_{jj} = \big(a_{ij}d_{jj}\big)b_{ij} = f_{ij}b_{ij},$$

donde

$$f_{ij} \ = \ \sum_{k=1}^{n} a_{ik}d_{kj},$$

luego, $E = (AD) \odot B$. Si se agrupan los términos de e_{ij} de otra manera, se tiene

$$e_{ij} \ = a_{ij}b_{ij}d_{jj} = a_{ij}\big(b_{ij}d_{jj}\big) = a_{ij}g_{ij},$$

donde

$$g_{ij} \ = \sum_{k=1}^{n} b_{ik}d_{kj},$$

y en este caso, $E = A \odot (BD)$.

9. $-$ 11. Quedan como ejercicio para el lector.

12. Supongamos que la matriz tiene la siguiente forma :

$$A = \begin{bmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{bmatrix}.$$

Entonces, el ángulo θ de rotación satisface las expresiones dadas en (5.62). Luego, A se puede expresar como

$$A = \frac{\lambda_1 - \lambda_2}{2} \begin{bmatrix} \frac{\lambda_1+\lambda_2}{\lambda_1-\lambda_2} + \cos 2\theta & \operatorname{sen} 2\theta \\ \operatorname{sen} 2\theta & \frac{\lambda_1+\lambda_2}{\lambda_1-\lambda_2} - \cos 2\theta \end{bmatrix}.$$

Por lo tanto, el producto Hadamard de A por ella misma k-veces sería

$$\bigodot_{j=1}^{k} A = \Big(\frac{\lambda_1 - \lambda_2}{2}\Big)^k \begin{bmatrix} \Big(\frac{\lambda_1+\lambda_2}{\lambda_1-\lambda_2} + \cos 2\theta\Big)^k & \big(\operatorname{sen} 2\theta\big)^k \\ \big(\operatorname{sen} 2\theta\big)^k & \Big(\frac{\lambda_1+\lambda_2}{\lambda_1-\lambda_2} - \cos 2\theta\Big)^k \end{bmatrix}.$$

■

Teorema 8.24 *Si A y B son matrices semidefinidas positivas de tamaño $n \times n$, entonces $A \odot B$ también lo es.*

Demostración.

Véase en Ding & Engle (2001). ■

Teorema 8.25 *Sean $A_1, A_2, \ldots, A_m$ una colección de m matrices no singulares de tamaño 2×2 y $A = \sum_{i=1}^{m} \alpha_i A_i$, con $\alpha_i \in \mathbb{R}$. Si $\det(A) \neq 0$, entonces*

$$A^{-1} = \sum_{i=1}^{m} \alpha_i \frac{\det\big(A_i\big)}{\det\big(A\big)} A_i^{-1}.$$

Demostración.

Al emplear el Teorema 3.13, se tiene que

$$A^2 - \operatorname{tr}\big(A\big)A + \det(A)I = \mathrm{O}, \tag{8.11}$$

y como $A = \sum_{i=1}^{m} \alpha_i A_i$, entonces

$$\operatorname{tr}(A) = \sum_{i=1}^{m} \alpha_i \operatorname{tr}\big(A_i\big).$$

Al sustituir en (8.11), se llega a

$$\left[\sum_{i=1}^{m} \alpha_i\, A_i\right]^2 - \left[\sum_{i=1}^{m} \alpha_i \operatorname{tr}\big(A_i\big)\right]\left[\sum_{i=1}^{m} \alpha_i\, A_i\right] + \det(A) I = \mathrm{O},$$

esta expresión se puede reescribir como sigue:

$$\begin{aligned}\sum_{i=1}^{m} \alpha_i^2 A_i^2 + \sum_{i=1}^{m}\sum_{j\neq i}^{m} \alpha_i\alpha_j\, A_i A_j - \sum_{i=1}^{m} \alpha_i^2 \operatorname{tr}\big(A_i\big) A_i \\ - \sum_{i=1}^{m}\sum_{j\neq i}^{m} \alpha_i\alpha_j \operatorname{tr}\big(A_j\big) A_i + \det(A) I = \mathrm{O},\end{aligned}$$

y agrupando términos, se obtiene

$$\sum_{i=1}^{m} \alpha_i^2 \big[A_i^2 - \operatorname{tr}\big(A_i\big) A_i\big] + \sum_{i=1}^{m}\sum_{j\neq i}^{m} \alpha_i\alpha_j\, A_i\big[A_j - \operatorname{tr}\big(A_j\big) I\big] + \det(A) I = \mathrm{O},$$

pero como cada una de las matrices A_i satisface (8.11), se tiene que

$$\sum_{i=1}^{m} \alpha_i^2 \big[-\det\big(A_i\big) I\big] + \sum_{i=1}^{m}\sum_{j\neq i}^{m} \alpha_i\alpha_j\, A_i A_j^{-1}\big[-\det\big(A_j\big) I\big] + \det(A) I = \mathrm{O}\,.$$

Aquí, se usó el hecho de que cada A_i es no singular, y reagrupando, se llega a

$$\begin{aligned}\det(A) I &= \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i\alpha_j \det\big(A_j\big) A_i A_j^{-1} \\ &= \sum_{i=1}^{m} \alpha_i\, A_i \sum_{j=1}^{m} \alpha_j \det\big(A_j\big) A_j^{-1} \\ &= A \sum_{j=1}^{m} \alpha_j \det\big(A_j\big) A_j^{-1},\end{aligned}$$

como por hipótesis $\det(A) \neq 0$, entonces A es invertible y, por lo tanto,

$$\det(A) A^{-1} = \sum_{j=1}^{m} \alpha_j \det\big(A_j\big) A_j^{-1}.$$

Al dividir esta última expresión por $\det(A)$, se completa la prueba. ■

Ejercicios 8.2

1. Encuentre el rango de $A \otimes B$, donde

$$A = \begin{bmatrix} 2 & 6 \\ 1 & 4 \\ 3 & 1 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 5 & 2 & 4 \\ 2 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

2. Sean A y B matrices de tamaño 2×2, dadas por:

$$A = \begin{bmatrix} 2 & 6 \\ 1 & 4 \end{bmatrix} \qquad \text{y} \qquad B = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$$

   Determine:

   a. $A \otimes B$, *b.* $B \otimes A$,

   c. $\operatorname{tr}\big(A \otimes B\big)$, *d.* $\det\big(A \otimes B\big)$,

   e. $\big(A \otimes B\big)^{-1}$, *f.* Los valores propios de $A \otimes B$.

3. Con las matrices del problema anterior, determine lo mismo pero para el producto Hadamard.

4. Sean A, B y C matrices de tamaño $m \times n$. Demuestre que

$$\operatorname{tr}\big[\big(A^t \odot B^t\big)C\big] = \operatorname{tr}\big[A^t\big(B \odot C\big)\big].$$

5. Sean A y B matrices simétricas semidefinidas positivas de tamaño $m \times m$. Demuestre que

   a. $\det\big(A \odot B\big) \geq \det\big(A\big)\det\big(B\big)$,

   b. $\det\big(A \odot A^{-1}\big) \geq 1$, si A es definida positiva.

6. Sean A y B matrices de tamaño $m \times m$ y $n \times n$, respectivamente, demuestre que:

$$\|A \otimes B\|_2 = \|A\|_2 \|B\|_2.$$

7. Sean A y B matrices de tamaño $n \times n$. Demuestre que para todo $k \in \mathbb{Z}$:

$$\big(I_n \otimes A\big)^k = I_n \otimes A^k \qquad \text{y} \qquad \big(B \otimes I_n\big)^k = B^k \otimes I_n.$$

Capítulo 9

Inversa generalizada de matrices

El concepto de inversa generalizada tiene sus principios en la teoría de ecuaciones lineales simultáneas (sistemas de m ecuaciones lineales con n incógnitas). La solución de un conjunto de ecuaciones lineales consistente

$$A\vec{x} = \vec{b}, \tag{9.1}$$

donde A es de tamaño $m \times n$ con rango $r \leq \text{mín}\{m, n\}$, puede asumir dos formas diferentes. Si $m = n = r$, el sistema (9.1) tiene solución única $\vec{x} = A^{-1}\vec{b}$. Sin embargo, cuando A es una matriz rectangular o singular, una representación simple de una solución en términos de A es más difícil. En este capítulo se tratarán estos sistemas de ecuaciones usando las inversas generalizadas de matrices. Dichas matrices las estudiaremos como una aplicación de las descomposiciones de matrices.

9.1 Definición y propiedades básicas

En esta sección, se analizarán las inversas generalizadas de matrices rectangulares o singulares. Estas inversas las estudiaremos como una aplicación de los valores propios, considerando los dos casos: valores propios reales o complejos.

Definición 9.1 Inversa generalizada ($\mathcal{IG}$)

Para cualquier matriz A cuadrada o rectangular, se dice que G es una inversa generalizada de A, si satisface las siguientes condiciones:

$$\begin{array}{ll} (i)\ AGA = A, & (iii)\ AG \text{ es simétrica e idempotente,} \\ (ii)\ GAG = G, & (iv)\ GA \text{ es simétrica e idempotente.} \end{array} \tag{9.2}$$

Nota 9.1 La inversa generalizada de A se llama también *seudoinversa* de A.

Teorema 9.1 *Si A es una matriz no singular, entonces $G = A^{-1}$.*

Demostración.
Queda como ejercicio para el lector. ■

Notación.
La notación y nomenclatura que se usará en este capítulo para los cuatro tipos de inversa generalizada introducido en (9.2), es el siguiente:

Condiciones que satisface	**Nombre**	**Abreviación**	**Notación**
(i)	$\mathcal{IG}$ condicionada	g_1-inversa	A^{g_1} o A^c
(i) y (ii)	$\mathcal{IG}$ reflexiva	g_2-inversa	A^{g_2} o A^r
$(i), (ii)$ y (iii)	$\mathcal{IG}$ normalizada	g_3-inversa	A^{g_3} o A^n
$(i), (ii)$ y (iv)	$\mathcal{IG}$ normalizada	g_3^*-inversa	$A^{g_3^*}$ o A^{n^*}
$(i), (ii), (iii)$ y (iv)	La $\mathcal{IG}$	g-inversa	A^g o A^-

Como veremos, el término "normalizada" significa de *norma mínima*.

En la Definición 9.1, no se establece que toda matriz tenga inversa generalizada y que además esta sea única. Por supuesto que así es, como lo establece el siguiente teorema.

Teorema 9.2 *Sea A una matriz cuadrada o rectangular, entonces:*

(1) *Siempre existe G.* (2) *G es única.*

Demostración.

(1) Si A es la matriz nula de tamaño $m \times n$, es claro que la g-inversa de A es la matriz nula de tamaño $n \times m$.

Si se supone que $\rho\big(A\big) = r > 0$, entonces por la propiedad (iv) del rango de una matriz (ver Capítulo 1), se tiene que existen K y L de tamaño $m \times r$ y $r \times n$, respectivamente, y ambas con rango r tales que

$$A = KL.$$

Entonces, la matriz dada por

$$A^g = L^t\big(LL^t\big)^{-1}\big(K^tK\big)^{-1}K^t \tag{9.3}$$

es una g-inversa de A, y al sustituir en (9.2), se obtiene que

(i) $AA^gA = KLL^t\big(LL^t\big)^{-1}\big(K^tK\big)^{-1}K^tKL = KL = A.$

(ii) $A^gAA^g = L^t\big(LL^t\big)^{-1}\big(K^tK\big)^{-1}K^t = A^g.$

(iii) $AA^g = KLL^t\big(LL^t\big)^{-1}\big(K^tK\big)^{-1}K^t = K\big(K^tK\big)^{-1}K^t.$

(iv) $A^gA = L^t\big(LL^t\big)^{-1}\big(K^tK\big)^{-1}K^tKL = L^t\big(LL^t\big)^{-1}L.$

Nótese que las matrices AA^g y A^gA son simétricas e idempotentes. Así pues, siempre existe una g-inversa de cualquier matriz A.

(2) Para probar la unicidad se procede por contradicción. Para ello, se supone que existen dos matrices A^g y B^g de tamaño $n \times m$, ambas inversas generalizadas de A.

Por ser A^g una g-inversa de A, se tiene que

$$AA^gA = A. \tag{9.4}$$

Al multiplicar por B^g, se obtiene

$$AA^gAB^g = AB^g,$$

y dada la simetría de AB^g y AA^g, resulta

$$AB^g = \big(AB^g\big)^t = \big[\big(AA^g\big)\big(AB^g\big)\big]^t = \big(AB^gA\big)A^g = AA^g. \tag{9.5}$$

De manera análoga, premultiplicando a (9.4) por B^g se llega a

$$B^gA = \big(B^gA\big)^t = \big[\big(B^gA\big)\big(A^gA\big)\big]^t = A^g\big(AB^gA\big) = A^gA. \tag{9.6}$$

Por último, si se premultiplica en (9.5) por B^g, se tiene

$$B^gAB^g = B^gAA^g,$$

y de acuerdo con (9.6) y la Definición 9.1, resulta

$$B^g = B^gAB^g = \left(B^gA\right)A^g = A^gAA^g = A^g.$$

Es decir, la g-inversa de una matriz es única. ■

Ejemplo 9.1

Dada la matriz $A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix}$, *determine qué tipo de inversa generalizada es* $G = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ 1 & -2 \\ 0 & 0 \end{bmatrix}$.

Solución.

Veamos las condiciones que cumple G de las dadas en (9.2):

$$AG = \frac{1}{3}\begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 1 & -2 \\ 0 & 0 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}. \tag{9.7}$$

Nótese que $AG = I$, la cual es simétrica e idempotente, luego,

$$AGA = \begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix} = A.$$

Por lo tanto, la matriz G es A^{g_1}. Veamos si cumple la segunda condición:

$$GA = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ 1 & -2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 3 & 0 & 3 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix}, \tag{9.8}$$

luego, GA no es simétrica, pero sí es idempotente. Por otra parte

$$GAG = \frac{1}{9}\begin{bmatrix} 6 & -3 \\ 3 & -6 \\ 0 & 0 \end{bmatrix} = G.$$

Así, G es una matriz A^{g_2} y de la expresión (9.7), se tiene finalmente que G es una A^{g_3}. No alcanza ser A^g, ya que no cumple la cuarta condición.

9.2 Propiedades de las inversas generalizadas

Algunas de las propiedades más importantes de la inversa generalizada se resumen en el siguiente teorema.

Teorema 9.3 *Sea A una matriz de tamaño $m \times n$ con rango $r \leq$ mín$\{m, n\}$ y A^g una matriz de tamaño $n \times m$. Entonces:*

a) $(A^g)^g = A$.

b) $(A^t)^g = (A^g)^t$.

c) $\big(AB\big)^g = B^g A^g$.

d) $A = AA^t\big(A^g\big)^t = \big(A^g\big)^t A^t A$.

e) $A^g = \big(A^t A\big)^g A^t = A^t\big(AA^t\big)^g$.

f) $\big(\alpha A\big)^g = \alpha^g A^g$, *donde* $\alpha \neq 0$ *es cualquier escalar con* $\alpha^g = \alpha^{-1}$.

g) Las matrices $I_m - AA^g$ *e* $I_n - A^g A$ *son idempotentes con rangos iguales a* $m - r$ *y* $n - r$, *respectivamente.*

h) $\rho(A^g) = \rho(A)$.

Demostración.

En esta demostración, se utilizan las condiciones dadas en (9.2).

a) Se tiene inmediatamente de las condiciones.

b) Supongamos que la g-inversa de A^t es $(A^t)^g$. Si se transpone la primera condición de la g-inversa de la matriz A, se tiene

$$\begin{aligned} [AA^g A]^t &= A^t \\ A^t (A^g)^t A^t &= A^t. \end{aligned}$$

Según el Teorema 9.2, la g-inversa es única, luego $(A^t)^g = (A^g)^t$.

c) - *f*) Quedan como ejercicio para el lector.

g) Para verificar si las matrices son idempotentes, se eleva cada una de ellas al cuadrado:

$$\begin{aligned}\left(I_m - AA^g\right)^2 &= I_m - AA^g - AA^g + (AA^gA)A^g = I_m - AA^g,\\ \left(I_n - A^gA\right)^2 &= I_n - A^gA - A^gA + A^g(AA^gA) = I_n - A^gA.\end{aligned}$$

Como por el Teorema 8.13 el rango de una matriz simétrica e idempotente es igual a su traza, se tiene que

$$\begin{aligned}\rho(I_m - AA^g) &= \operatorname{tr}(I_m - AA^g) = m - \operatorname{tr}(AA^g)\\ &= m - \rho(AA^g) = m - r,\\ \rho(I_n - A^gA) &= \operatorname{tr}(I_n - A^gA) = n - \operatorname{tr}(A^gA)\\ &= n - \rho(A^gA) = n - r.\end{aligned}$$

Aquí se uso el hecho que el rango de un producto no puede exceder el rango más pequeño de los factores, es decir

$$\begin{aligned}r = \rho(A) &= \rho(AA^gA) \leq \rho(AA^g) \leq \rho(A) = r,\\ r = \rho(A) &= \rho(AA^gA) \leq \rho(A^gA) \leq \rho(A) = r.\end{aligned}$$

Por el teorema del emparedado, se tienen las igualdades.

h) Por la parte a), si $AA^gA = A$, entonces

$$\rho(A) = \rho(AA^gA) \leq \rho(AA^g) \leq \rho(A^g).$$

Por otra parte, la condición $A^gAA^g = A^g$ implica que

$$\rho(A^g) = \rho(A^gAA^g) \leq \rho(AA^g) \leq \rho(A).$$

Así, $\rho(A^g) = \rho(A)$. ■

Teorema 9.4 *Si A es una matriz simétrica, entonces A^g es simétrica.*

Demostración.

La prueba se sigue de la parte *b*) del Teorema 9.3, es decir,

$$A^g = \left(A^t\right)^g = \left(A^g\right)^t. \quad ■$$

Corolario 9.4.1 *Si A es simétrica e idempotente, entonces:*

$$A^g \neq A.$$

Demostración.
Queda como ejercicio para el lector. ■

Ejemplo 9.2 *Determine la inversa generalizada de la matriz asociada a la forma cuadrática del Ejemplo 5.5.*

Solución.
En el Ejemplo 8.4, se mostró que $I_n - \overline{J}_n$ era simétrica e idempotente. Luego, por el Corolario 9.4.1, se tiene que

$$\left(I_n - \overline{J}_n\right)^g = I_n - \overline{J}_n.$$

Teorema 9.5 *Si A y A^{g_3} son simétricas, entonces $A^{g_3} = A^g$.*

Demostración.
Puesto que A^{g_3} es simétrica,

$$\left(A^{g_3}A\right)^t = AA^{g_3} = \left(AA^{g_3}\right)^t = A^{g_3}A,$$

y la cuarta condición dada en (9.2) se satisface. ■

9.3 Métodos para calcular inversas generalizadas

En esta sección, se ilustran algunos de los métodos para hallar la g-inversa. Se desarrollan solo los métodos que utilizan las distintas factorizaciones de la matriz A estudiadas en este material.

Aunque en esta sección se consideran únicamente matrices reales, cuando el lector necesite emplear alguno de los métodos desarrollados aquí para matrices complejas, simplemente puede realizar los cambios adecuados en cada método. Por ejemplo, en vez de utilizar A^t se usa A^H y si en el método se emplea una matriz ortogonal, pues se cambia por una matriz unitaria.

Teorema 9.6 *Sea A una matriz real de tamaño $m \times n$ con rango $r \leq$ mín$\{m, n\}$ particionada como:*

$$A = \begin{bmatrix} A_{11} & \vdots & A_{12} \\ \dots & . & \dots \\ A_{21} & \vdots & A_{22} \end{bmatrix},$$

donde A_{11} ó A_{22} es una submatriz de tamaño $r \times r$. Entonces:

1. *Si A_{11} es no singular y $\left(A/A_{11}\right) = \mathrm{O}$, una g_2-inversa de A es la matriz A^{g_2} de tamaño $n \times m$ dada por*

$$A^{g_2} = \begin{bmatrix} A_{11}^{-1} & \vdots & \mathrm{O}_{r \times m_1} \\ \dots & . & \dots \\ \mathrm{O}_{n_1 \times r} & \vdots & \mathrm{O}_{n_1 \times m_1} \end{bmatrix}. \tag{9.9}$$

   con $m_1 = m - r$ y $n_1 = n - r$.

2. *Si A_{22} es no singular y $\left(A/A_{22}\right) = \mathrm{O}$, una g_2-inversa de A es la matriz A^{g_2} de tamaño $n \times m$ dada por*

$$A^{g_2} = \begin{bmatrix} \mathrm{O}_{n_1 \times m_1} & \vdots & \mathrm{O}_{n_1 \times r} \\ \dots & . & \dots \\ \mathrm{O}_{r \times m_1} & \vdots & A_{22}^{-1} \end{bmatrix}. \tag{9.10}$$

Demostración.

Como la partición de las matrices expuestas son consistentes para el producto, efectúe los productos $AA^{g_2}A$ y $A^{g_2}AA^{g_2}$. Obsérvese que se obtiene respectivamente A y A^{g_2}. ■

Ejemplo 9.3

Dada la matriz $A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix}$, *obtenga una g_2-inversa.*

Solución.

La partición de tamaño $1 \times (2+1)$ de la matriz dada es

$$A = \left[\begin{array}{cc:c} 2 & -1 & 3 \\ 1 & -2 & 3 \end{array}\right] = (A_{11} \vdots A_{12}).$$

Al calcular la inversa de A_{11}, se obtiene

$$A_{11}^{-1} = -\frac{1}{3}\begin{bmatrix} -2 & 1 \\ -1 & 2 \end{bmatrix}.$$

Luego, una g_2-inversa es la matriz

$$A^{g_2} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ 1 & -2 \\ 0 & 0 \end{bmatrix}.$$

Nótese que esta es igual a la dada en el Ejemplo 9.1.

Corolario 9.6.1 *Sea A una matriz "diagonal" de tamaño $m \times n$ y rango $r \leq \min\{m,n\}$ particionada como sigue:*

$$A = \left[\begin{array}{c:c:c} D_{r\times r} & \vdots & \mathrm{O}_{r\times n_1} \\ \hdashline \cdots & \cdot & \cdots \\ \hdashline \mathrm{O}_{m_1\times r} & \vdots & \mathrm{O}_{m_1\times n_1} \end{array}\right], \quad \textit{donde} \quad D_{r\times r} = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_r \end{bmatrix}$$

con $m_1 = m - r$ y $n_1 = n - r$. Entonces, la inversa generalizada de A está dada por

$$G = \left[\begin{array}{c:c} D_{r\times r}^{-1} & \vdots & \mathrm{O}_{r\times m_1} \\ \hdashline \cdots & & \cdots \\ \hdashline \mathrm{O}_{n_1\times r} & \vdots & \mathrm{O}_{n_1\times m_1} \end{array}\right]. \tag{9.11}$$

Demostración.

Queda como ejercicio para el lector. ■

Teorema 9.7 *Sea A una matriz de tamaño $n \times n$ y rango m $(m < n)$ con valores propios no nulos $\lambda_1, \lambda_2, \ldots, \lambda_m$. Si A se puede factorizar como en* (2.33)*, entonces la A^{g_2} viene dada por*

$$A^{g_2} = \sum_{i=1}^{m} \frac{1}{\lambda_k} \mathbf{E}(\lambda_k), \tag{9.12}$$

donde $\mathbf{E}(\lambda_k)$ está dada en (2.11)*.*

Demostración.

Puesto que A se puede factorizar como en (2.33), veamos si $G = A^{g_2}$ cumple la primera condición de la Definición 9.1:

$$\begin{aligned} AGA &= A \left(\sum_{i=1}^{m} \frac{1}{\lambda_k} \mathbf{E}(\lambda_k) \right) A = A \sum_{i=1}^{m} \frac{1}{\lambda_k} \big(\mathbf{E}(\lambda_k) A \big) \\ &= A \sum_{i=1}^{m} \mathbf{E}(\lambda_k) = A. \end{aligned}$$

Luego, G es una matriz A^{g_1}. Veamos si es A^{g_2}.

$$\begin{aligned} GAG &= \left(\sum_{i=1}^{m} \frac{1}{\lambda_k} \mathbf{E}(\lambda_k) \right) A \left(\sum_{i=1}^{m} \frac{1}{\lambda_k} \mathbf{E}(\lambda_k) \right) \\ &= \left(\sum_{i=1}^{m} \mathbf{E}(\lambda_k) \right) \left(\sum_{i=1}^{m} \frac{1}{\lambda_k} \mathbf{E}(\lambda_k) \right) = G. \end{aligned}$$

Aquí se emplearon las propiedades dadas para las matrices $\mathbf{E}(\lambda_k)$ en el Teorema 2.12. Ahora, probemos si G es una matriz A^{g_3}:

$$AG = A \left(\sum_{i=1}^{m} \frac{1}{\lambda_k} \mathbf{E}(\lambda_k) \right) = \sum_{i=1}^{m} \frac{1}{\lambda_k} \big(A \mathbf{E}(\lambda_k) \big) = \sum_{i=1}^{m} \mathbf{E}(\lambda_k).$$

Nótese que la última matriz no es simétrica, pero sí es idempotente. De manera análoga, se obtiene que GA es una matriz idempotente pero no es simétrica, luego G es una g_2-inversa de A. ■

Ejemplo 9.4 $\mathcal{IG}$ **usando la descomposición de Sylvester**

Sea $A = \begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix}$*, obtenga una g_2-inversa.*

Solución.

En este caso, los valores propios de A son $\lambda_1 = 7$, $\lambda_2 = 3$ y $\lambda_3 = 0$, y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 7 \\ 3 \\ -5 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 1 \\ -6 \\ 1 \end{bmatrix}$, respectivamente. Por otra parte, los vectores propios correspondientes de A^t son $\vec{w}_1 = \begin{bmatrix} 9 \\ 4 \\ 15 \end{bmatrix}$, $\vec{w}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ y $\vec{w}_3 = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$, respectivamente. Luego, las matrices de proyección espectral $\mathbf{E}(\lambda_k)$ son

$$\mathbf{E}(\lambda_1) = \frac{\vec{v}_1\vec{w}_1^t}{\vec{w}_1^t\vec{v}_1} = \frac{1}{28}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} 9 & 4 & 15 \end{bmatrix} = \frac{1}{28}\begin{bmatrix} 9 & 4 & 15 \\ 9 & 4 & 15 \\ 9 & 4 & 15 \end{bmatrix}$$

$$\mathbf{E}(\lambda_2) = \frac{\vec{v}_2\vec{w}_2^t}{\vec{w}_2^t\vec{v}_2} = \frac{1}{12}\begin{bmatrix} 7 \\ 3 \\ -5 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 \end{bmatrix} = \frac{1}{12}\begin{bmatrix} 7 & 0 & -7 \\ 3 & 0 & -3 \\ -5 & 0 & 5 \end{bmatrix}$$

$$\mathbf{E}(\lambda_3) = \frac{\vec{v}_2\vec{w}_2^t}{\vec{w}_2^t\vec{v}_2} = \frac{1}{21}\begin{bmatrix} 1 \\ -6 \\ 1 \end{bmatrix}\begin{bmatrix} 2 & -3 & 1 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 2 & -3 & 1 \\ -12 & 18 & -6 \\ 2 & -3 & 1 \end{bmatrix}.$$

Al sustituir en (9.12), se llega a

$$G = \frac{1}{441}\begin{bmatrix} 106 & 9 & -52 \\ 57 & 9 & -3 \\ -41 & 9 & 95 \end{bmatrix},$$

de manera que

$$AG = \frac{1}{441}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix}\begin{bmatrix} 106 & 9 & -52 \\ 57 & 9 & -3 \\ -41 & 9 & 95 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 19 & 3 & -1 \\ 12 & 3 & 6 \\ -2 & 3 & 20 \end{bmatrix},$$

$$GA = \frac{1}{441}\begin{bmatrix} 106 & 9 & -52 \\ 57 & 9 & -3 \\ -41 & 9 & 95 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 19 & 3 & -1 \\ 12 & 3 & 6 \\ -2 & 3 & 20 \end{bmatrix}.$$

Entonces $AG = GA$, pero los productos no dan como resultado matrices simétricas aunque son idempotentes. Por otra parte,

$$AGA = \frac{1}{21}\begin{bmatrix} 19 & 3 & -1 \\ 12 & 3 & 6 \\ -2 & 3 & 20 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 84 & 21 & 42 \\ 63 & 21 & 63 \\ 21 & 21 & 105 \end{bmatrix} = A$$

y

$$GAG = \frac{1}{9261}\begin{bmatrix} 19 & 3 & -1 \\ 12 & 3 & 6 \\ -2 & 3 & 20 \end{bmatrix}\begin{bmatrix} 106 & 9 & -52 \\ 57 & 9 & -3 \\ -41 & 9 & 95 \end{bmatrix} = \frac{3}{9261}\begin{bmatrix} 742 & 63 & -364 \\ 399 & 63 & -21 \\ -287 & 63 & 665 \end{bmatrix}.$$

Al simplificar, se obtiene la matriz G, la cual satisface las condiciones (i) y (ii) dadas en (9.2), pero no (iii) y (iv), pues AG y GA no son matrices simétricas aunque son idempotentes.

Teorema 9.8 *Sea A una matriz simétrica de tamaño $n \times n$ y rango r, $(r < n)$. Entonces, la inversa generalizada de A está dada por*

$$G = P\Lambda^g P^{-1}, \tag{9.13}$$

donde P es una matriz real de tamaño $n \times n$ cuyas columnas son los vectores propios asociados a A, particionada como:

$$P = [S \vdots T] = [\vec{v}_1 \ \ldots \ \vec{v}_r \vdots \vec{v}_{r+1} \ \ldots \ \vec{v}_n]. \tag{9.14}$$

Aquí, la submatriz S es de tamaño $n \times r$, sus columnas corresponden a los vectores propios asociados a los valores propios distintos de cero de la matriz A y la submatriz T es de tamaño $n \times (n - r)$, cuyas columnas corresponden a los vectores propios asociados a los valores propios nulos de A.

$$\Lambda^g = \begin{bmatrix} D^{-1} & \mathrm{O}^* \\ \mathrm{O}^* & \mathrm{O} \end{bmatrix}, \tag{9.15}$$

con D la submatriz real de tamaño $r \times r$ que tiene en la diagonal los valores propios distintos de cero asociados a A, y con O *la submatriz real de tamaño $(n-r) \times (n-r)$ en cuya diagonal están los valores propios nulos de A.*

Demostración.

Puesto que A tiene n vectores propios $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ que corresponden a los valores propios (no necesariamente diferentes) $\lambda_1, \lambda_2, \ldots, \lambda_n$, dichos vectores resultan ser linealmente independientes, y por lo tanto, la matriz P dada en (9.14) es no singular.

Por consiguiente, la matriz A se puede expresar como $A = P\Lambda P^{-1}$, donde $\Lambda = \begin{bmatrix} D & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}$.

Veamos si $G = P\Lambda^g P^{-1}$ cumple la primera condición de la Definición 9.1:

$$\begin{aligned} AGA &= A\big(P\Lambda^g P^{-1}\big)A = \big(P\Lambda P^{-1}\big)\big(P\Lambda^g P^{-1}\big)\big(P\Lambda P^{-1}\big) \\ &= P\Lambda\Lambda^g\Lambda P^{-1} = P\Lambda P^{-1} = A. \end{aligned}$$

Luego, G es una matriz A^{g_1}. Observemos si es A^{g_2}:

$$\begin{aligned} GAG &= \big(P\Lambda^g P^{-1}\big)A\big(P\Lambda^g P^{-1}\big) = \big(P\Lambda^g P^{-1}\big)\big(P\Lambda P^{-1}\big)\big(P\Lambda^g P^{-1}\big) \\ &= P\Lambda^g\Lambda\Lambda^g P^{-1} = P\Lambda^g P^{-1} = G. \end{aligned}$$

Ahora, verifiquemos si G es una matriz A^{g_3}:

$$AG = A\big(P\Lambda^g P^{-1}\big) = \big(P\Lambda P^{-1}\big)(P\Lambda^g P^{-1}) = P\big(\Lambda\Lambda^g\big)P^{-1}. \qquad (9.16)$$

Pero como $A = A^t$, por el Teorema 2.29, la matriz A es semejante a una matriz Q ortogonal. Si se ortonormalizan las columnas de la matriz P, se tiene que $P^{-1} = P^t$ y, por lo tanto,

$$(AG)^t = [P(\Lambda\Lambda^g)P^{-1}]^t = \left\{P\begin{bmatrix} I_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}P^t\right\}^t = P\begin{bmatrix} I_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}P^t = AG.$$

También, G es A^{g_3}. Observemos si cumple la cuarta condición dada en (9.2)

$$GA = (P\Lambda^g P^{-1})A = (P\Lambda^g P^{-1})(P\Lambda P^{-1}) = P(\Lambda^g\Lambda)P^{-1}. \qquad (9.17)$$

Usando de nuevo el hecho de que A es diagonalizable ortogonalmente, se tiene

$$(GA)^t = [P(\Lambda^g\Lambda)P^{-1}]^t = \left\{P\begin{bmatrix} I_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O}\end{bmatrix}P^t\right\}^t = P\begin{bmatrix} I_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O}\end{bmatrix}P^t = GA.$$

Así, G es la g-inversa de A y el teorema queda demostrado. ■

Ejemplo 9.5 Inversa generalizada de una matriz simétrica

Sea $A = \begin{bmatrix} 5 & -4 & 9 \\ -4 & 5 & -9 \\ 9 & -9 & 18 \end{bmatrix}$, *obtenga la g-inversa.*

Solución.

En este caso, la ecuación característica es:

$$\det(A - \lambda I) = -\lambda^3 + 28\lambda^2 - 27\lambda = 0.$$

Entonces, los valores propios de A son $\lambda_1 = 1$, $\lambda_2 = 27$ y $\lambda_3 = 0$.

Para $\lambda_1 = 1$, se tiene el vector propio correspondiente $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

Si $\lambda_2 = 27$, se obtiene el vector propio asociado $\vec{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

y para $\lambda_3 = 0$, se llega al vector propio $\vec{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$.

Estableciendo

$$P = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 0 & 2 & 1 \end{bmatrix} \qquad \text{con} \qquad P^{-1} = -\frac{1}{6}\begin{bmatrix} -3 & -3 & 0 \\ -1 & 1 & -2 \\ 2 & -2 & -2 \end{bmatrix}$$

y

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 27 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{con} \qquad \Lambda^g = \frac{1}{27}\begin{bmatrix} 27 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

se obtiene

$$G = -\frac{1}{162}\begin{bmatrix}1 & 1 & -1\\ 1 & -1 & 1\\ 0 & 2 & 1\end{bmatrix}\begin{bmatrix}27 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}-3 & -3 & 0\\ -1 & 1 & -2\\ 2 & -2 & -2\end{bmatrix}.$$

Después de realizar la multiplicación de las matrices, queda

$$G = \frac{1}{81}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix},$$

de manera que

$$AG = \frac{1}{81}\begin{bmatrix}5 & -4 & 9\\ -4 & 5 & -9\\ 9 & -9 & 18\end{bmatrix}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix} = \frac{1}{3}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{bmatrix},$$
$$GA = \frac{1}{81}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix}\begin{bmatrix}5 & -4 & 9\\ -4 & 5 & -9\\ 9 & -9 & 18\end{bmatrix} = \frac{1}{3}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{bmatrix}.$$

Así, $AG = GA$ y, además, los productos dan como resultado matrices simétricas e idempotentes. Por otra parte,

$$AGA = \frac{1}{3}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{bmatrix}\begin{bmatrix}5 & -4 & 9\\ -4 & 5 & -9\\ 9 & -9 & 18\end{bmatrix} = \frac{1}{3}\begin{bmatrix}15 & -12 & 27\\ -12 & 15 & -27\\ 27 & -27 & 54\end{bmatrix} = A,$$
$$GAG = \frac{1}{243}\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & -1\\ 1 & -1 & 2\end{bmatrix}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix} = \frac{3}{243}\begin{bmatrix}41 & 40 & 1\\ 40 & 41 & -1\\ 1 & -1 & 2\end{bmatrix} = G.$$

Corolario 9.8.1 *Sea A una matriz singular de tamaño $n \times n$ con valores propios (reales o complejos) distintos de cero $\lambda_1, \lambda_2, \ldots, \lambda_r$ ($r = \rho(A)$). Entonces, una g_2-inversa de A es la matriz definida de la siguiente forma:*

$$A^{g_2} = P\Lambda^g P^{-1}, \tag{9.18}$$

donde P^{1} *y* Λ^{g} *están definidas de manera análoga a* (9.14) *y* (9.15), *y*

$$D = \begin{cases} \mathcal{J} & \textit{si } m.g.(\lambda_i) \leq m.a.(\lambda_i) \\ R & \textit{si posee valores propios complejos,} \end{cases}$$

en donde $\mathcal{J}$ *es la matriz de Jordan dada en* (3.11) *y la matriz* R, *en* (2.23). *En este caso,* $AG = GA$.

Demostración.

En el Teorema 9.8, se demostró que G era A^{g_2}. Para demostrar que $AG = GA$, de (9.16) y (9.17), se tiene que

$$AG = P(\Lambda\Lambda^{g})P^{-1} \qquad \text{y} \qquad GA = P(\Lambda^{g}\Lambda)P^{-1}.$$

Dado que $\Lambda\Lambda^{g} = \Lambda^{g}\Lambda = \begin{bmatrix} I_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}$, el corolario queda demostrado. ■

Ejemplo 9.6 $\mathcal{IG}$ **de una matriz con valores propios reales**

Obtenga una g_2*-inversa mediante el método propuesto para la matriz dada en el Ejemplo 9.4.*

Solución.

En el Ejemplo 9.4, se obtuvo que los valores propios de A eran $\lambda_1 = 7$, $\lambda_2 = 3$ y $\lambda_3 = 0$, y los vectores propios correspondientes eran $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} -7 \\ -3 \\ 5 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 1 \\ -6 \\ 1 \end{bmatrix}$, respectivamente. Estableciendo

$$P = \begin{bmatrix} 1 & -7 & 1 \\ 1 & -3 & -6 \\ 1 & 5 & 1 \end{bmatrix} \qquad \text{con} \qquad P^{-1} = \frac{1}{84}\begin{bmatrix} 27 & 12 & 45 \\ -7 & 0 & 7 \\ 8 & -12 & 4 \end{bmatrix}$$

y

$$\Lambda = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{con} \qquad \Lambda^{g} = \frac{1}{21}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

[1] Cuando la multiplicidad algebraica de un λ_i sea mayor que su multiplicidad geométrica, algunas de las columnas de P serán vectores propios generalizados.

se obtiene

$$G = \frac{1}{1764}\begin{bmatrix} 1 & -7 & 1 \\ 1 & -3 & -6 \\ 1 & 5 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 27 & 12 & 45 \\ -7 & 0 & 7 \\ 8 & -12 & 4 \end{bmatrix}.$$

Después de realizar el producto entre matrices, se llega a

$$G = \frac{1}{441}\begin{bmatrix} 106 & 9 & -52 \\ 57 & 9 & -3 \\ -41 & 9 & 95 \end{bmatrix}.$$

Nótese que esta matriz coincide con la obtenida en el Ejemplo 9.4, luego la matriz G es una g_2-inversa de A.

Corolario 9.8.2 *Sea A una matriz real de tamaño $m \times n$ $(n < m)$ y rango r, $(r \leq n)$. Entonces, la g-inversa de A está dada por*

$$G = (A^tA)^gA^t, \tag{9.19}$$

donde $(A^tA)^g = P\Lambda^gP^{-1}$ es la matriz definida en (9.13)*. Si $r = n$, entonces $G = (A^tA)^{-1}A^t$ y $GA = I_n$.*

Demostración.

En el Teorema 9.8, se demostró que la g-inversa de matrices simétricas cumplen las condiciones establecidas en la Definición 9.1. Entonces, $(A^tA)^g$ las cumple. Veamos si la expresión dada en (9.19) verifica las condiciones dadas en (9.2):

$$AG = A(A^tA)^gA^t.$$

Pero por la propiedad d) del Teorema 9.3, se tiene que $A = (A^t)^gA^tA$, luego,

$$\begin{aligned} AGA &= [(A^t)^gA^tA](A^tA)^gA^tA = (A^t)^g(A^tA)(A^tA)^g(A^tA) \\ &= (A^t)^g(A^tA) = A. \end{aligned}$$

Por otra parte,

$$GA = (A^tA)^gA^tA.$$

Entonces,

$$\begin{aligned} GAG &= (A^tA)^gA^tA(A^tA)^gA^t = (A^tA)^g(A^tA)(A^tA)^gA^t \\ &= (A^tA)^gA^t = G. \end{aligned}$$

Ahora, observemos si AG y GA son simétricas e idempotentes:

$$\begin{aligned} (AG)^t &= [A(A^tA)^gA^t]^t = A[(A^tA)^g]^tA^t \\ &= A[(A^tA)^t]^gA^t = AG \end{aligned}$$

y

$$\begin{aligned} (GA)^t &= [(A^tA)^gA^tA]^t = [(A^tA)^g(A^tA)]^t \\ &= (A^tA)^g(A^tA) = GA; \end{aligned}$$

la última expresión se tiene debido a que $(A^tA)^g$ es una g-inversa de (A^tA). ■

Corolario 9.8.3 *Sea A una matriz real de tamaño $m \times n$ $(m < n)$ y rango r, $(r \leq m)$. Entonces, la g-inversa de A es*

$$G = A^t(AA^t)^g, \tag{9.20}$$

donde $(AA^t)^g = P\Lambda^gP^{-1}$ es la matriz definida en (9.13). *Si $r = m$, entonces $G = A^t(AA^t)^{-1}$ y $AG = I_m$.*

Demostración.
Queda como ejercicio para el lector. ■

Ejemplo 9.7 *Considérese la matriz dada en el Ejemplo 9.1, obtenga la g-inversa.*

Solución.
Como $\rho(A) = 2$, el producto de AA^t da como resultado

$$AA^t = \begin{bmatrix} 14 & 13 \\ 13 & 14 \end{bmatrix} \qquad \text{y} \qquad (AA^t)^{-1} = \frac{1}{27}\begin{bmatrix} 14 & -13 \\ -13 & 14 \end{bmatrix}.$$

Luego, la inversa generalizada es

$$A^g = A^t\left(AA^t\right)^{-1} = \frac{1}{9}\begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix}.$$

la cual es diferente a la A^{g_2} dada en el Ejemplo 9.1.

En el caso de que no se establezca primero el rango de la matriz A, se puede realizar el producto de A^tA, el cual da como resultado

$$A^tA = \begin{bmatrix} 5 & -4 & 9 \\ -4 & 5 & -9 \\ 9 & -9 & 18 \end{bmatrix}.$$

En el Ejemplo 9.5, se obtuvo que la g-inversa para esta matriz era

$$(A^tA)^g = \frac{1}{81}\begin{bmatrix} 41 & 40 & 1 \\ 40 & 41 & -1 \\ 1 & -1 & 2 \end{bmatrix}.$$

Por lo tanto, la g-inversa de la matriz A es

$$A^g = \frac{1}{9}\begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix},$$

la cual coincide con la obtenida anteriormente.

Ejemplo 9.8 *Determine una g-inversa para la matriz:*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix}.$$

Solución.

El producto de AA^t da como resultado:

$$B = AA^t = \begin{bmatrix} 14 & 7 & 7 \\ 7 & 6 & 1 \\ 7 & 1 & 6 \end{bmatrix}.$$

En este caso, los valores propios de B son $\lambda_1 = 21$, $\lambda_2 = 5$ y $\lambda_3 = 0$, y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$, respectivamente. Estableciendo

$$P = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \text{con} \qquad P^{-1} = \frac{1}{6} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -3 & 3 \\ -2 & 2 & 2 \end{bmatrix}$$

y

$$\Lambda = \begin{bmatrix} 21 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{con} \qquad \Lambda^g = \frac{1}{105} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 21 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

se obtiene

$$(AA^t)^g = \frac{1}{630} \begin{bmatrix} 2 & 0 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 21 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -3 & 3 \\ -2 & 2 & 2 \end{bmatrix}.$$

Después de multiplicar las matrices, queda

$$(AA^t)^g = \frac{1}{315} \begin{bmatrix} 10 & 5 & 5 \\ 5 & 34 & -29 \\ 5 & -29 & 34 \end{bmatrix}.$$

Por lo tanto, la g-inversa de la matriz A es

$$A^g = \frac{1}{105} \begin{bmatrix} 5 & -29 & 34 \\ 10 & 5 & 5 \\ 15 & 18 & -3 \end{bmatrix}.$$

El lector puede verificar que esta matriz cumple las condiciones dadas en (9.2).

Teorema 9.9 *Supóngase que $A = LU$ es una descomposición de la matriz A de tamaño $m \times n$, de rango $r \leq \min\{m, n\}$. Entonces, la inversa generalizada de A está dada por*

$$G = \widetilde{U}^t \big(\widetilde{U}\ \widetilde{U}^t\big)^{-1} \big(\widetilde{L}^t \widetilde{L}\big)^{-1} \widetilde{L}^t, \tag{9.21}$$

donde $\widetilde{U}$ es una matriz de tamaño $n \times r$ de rango r, obtenida de eliminar las filas nulas de U, y la matriz $\widetilde{L}$ de tamaño $m \times r$, también de rango r, es obtenida eliminando las columnas que multiplican a las respectivas filas nulas de la matriz U.

Demostración.
Queda como ejercicio para el lector. ■

Ejemplo 9.9 *Considérese la transpuesta de la matriz dada en el Ejemplo 9.1 y utilice el Teorema 9.9 para hallar la g-inversa.*

Solución.
Al transponer la matriz del Ejemplo 9.1, la factorización LU es

$$\begin{bmatrix} 2 & 1 \\ -1 & -2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ \frac{3}{2} & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & -\frac{3}{2} \\ 0 & 0 \end{bmatrix} = LU.$$

Si se elimina la última fila de U y la última columna de L, se obtiene

$$\begin{bmatrix} 2 & 1 \\ -1 & -2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \\ \frac{3}{2} & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & -\frac{3}{2} \end{bmatrix} = \widetilde{L}\,\widetilde{U}.$$

Luego,

$$\widetilde{U}\,\widetilde{U}^t = \begin{bmatrix} 2 & 1 \\ 0 & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & -\frac{3}{2} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 20 & -6 \\ -6 & 9 \end{bmatrix},$$

y por lo tanto

$$\widetilde{U}^t\big(\widetilde{U}\,\widetilde{U}^t\big)^{-1} = \begin{bmatrix} 2 & 0 \\ 1 & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{6} \\ \frac{1}{6} & \frac{5}{9} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 3 & 2 \\ 0 & -4 \end{bmatrix}.$$

Por otra parte,

$$\widetilde{L}^t\widetilde{L} = \begin{bmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \\ \frac{3}{2} & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 7 & -4 \\ -4 & 4 \end{bmatrix},$$

de donde

$$\left(\widetilde{L}^t\widetilde{L}\right)^{-1}\widetilde{L}^t = \frac{1}{6}\begin{bmatrix}4 & 4\\ 4 & 7\end{bmatrix}\begin{bmatrix}1 & -\frac{1}{2} & \frac{3}{2}\\ 0 & 1 & -1\end{bmatrix} = \frac{1}{6}\begin{bmatrix}4 & 2 & 2\\ 4 & 5 & -1\end{bmatrix}.$$

Siguiendo el procedimiento dado en (9.21), se tiene que

$$G = \frac{1}{36}\begin{bmatrix}3 & 2\\ 0 & -4\end{bmatrix}\begin{bmatrix}4 & 2 & 2\\ 4 & 5 & -1\end{bmatrix} = \frac{1}{9}\begin{bmatrix}5 & 4 & 1\\ -4 & -5 & 1\end{bmatrix},$$

la cual coincide con la transpuesta obtenida en el Ejemplo 9.7.

Ejemplo 9.10 *Considérese la matriz dada en el Ejemplo 9.4, obtenga la factorización LU de A y utilice el Teorema 9.9 para hallar la g-inversa.*

Solución.

La factorización LU de la matriz dada en el Ejemplo 9.4 es

$$\begin{bmatrix}4 & 1 & 2\\ 3 & 1 & 3\\ 1 & 1 & 5\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ \frac{3}{4} & 1 & 0\\ \frac{1}{4} & 3 & 1\end{bmatrix}\begin{bmatrix}4 & 1 & 2\\ 0 & \frac{1}{4} & \frac{3}{2}\\ 0 & 0 & 0\end{bmatrix} = LU.$$

Al eliminar la última fila de U y la última columna de L, se obtiene

$$\begin{bmatrix}4 & 1 & 2\\ 3 & 1 & 3\\ 1 & 1 & 5\end{bmatrix} = \begin{bmatrix}1 & 0\\ \frac{3}{4} & 1\\ \frac{1}{4} & 3\end{bmatrix}\begin{bmatrix}4 & 1 & 2\\ 0 & \frac{1}{4} & \frac{3}{2}\end{bmatrix} = \widetilde{L}\,\widetilde{U}.$$

Luego,

$$\widetilde{U}\,\widetilde{U}^t = \begin{bmatrix}4 & 1 & 2\\ 0 & \frac{1}{4} & \frac{3}{2}\end{bmatrix}\begin{bmatrix}4 & 0\\ 1 & \frac{1}{4}\\ 2 & \frac{3}{2}\end{bmatrix} = \frac{1}{16}\begin{bmatrix}336 & 52\\ 52 & 37\end{bmatrix},$$

y por lo tanto

$$\widetilde{U}^t\left(\widetilde{U}\,\widetilde{U}^t\right)^{-1} = \frac{1}{608}\begin{bmatrix}4 & 0\\ 1 & \frac{1}{4}\\ 2 & \frac{3}{2}\end{bmatrix}\begin{bmatrix}37 & -52\\ -52 & 336\end{bmatrix} = \frac{1}{152}\begin{bmatrix}37 & -52\\ 6 & 8\\ -1 & 100\end{bmatrix}.$$

Por otra parte,

$$\widetilde{L}^t\widetilde{L} = \begin{bmatrix}1 & \frac{3}{4} & \frac{1}{4}\\ 0 & 1 & 3\end{bmatrix}\begin{bmatrix}1 & 0\\ \frac{3}{4} & 1\\ \frac{1}{4} & 3\end{bmatrix} = \frac{1}{8}\begin{bmatrix}13 & 12\\ 12 & 80\end{bmatrix},$$

de donde

$$\left(\widetilde{L}^t\widetilde{L}\right)^{-1}\widetilde{L}^t = \frac{1}{112}\begin{bmatrix} 80 & -12 \\ -12 & 13 \end{bmatrix}\begin{bmatrix} 1 & \frac{3}{4} & \frac{1}{4} \\ 0 & 1 & 3 \end{bmatrix} = \frac{1}{28}\begin{bmatrix} 20 & 12 & -4 \\ -3 & 1 & 9 \end{bmatrix}.$$

Siguiendo el procedimiento dado en (9.21), se tiene que

$$G = \frac{1}{4256}\begin{bmatrix} 37 & -52 \\ 6 & 8 \\ -1 & 100 \end{bmatrix}\begin{bmatrix} 20 & 12 & -4 \\ -3 & 1 & 9 \end{bmatrix} = \frac{8}{4256}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix}.$$

Nótese que esta matriz no coincide con la obtenida en el Ejemplo 9.4, por otra parte

$$AG = \frac{1}{532}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 10 & 6 & -2 \\ 6 & 5 & 3 \\ -2 & 3 & 13 \end{bmatrix} \quad \text{y}$$

$$GA = \frac{1}{532}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix} = \frac{1}{38}\begin{bmatrix} 37 & 6 & -1 \\ 6 & 2 & 6 \\ -1 & 6 & 37 \end{bmatrix}.$$

En este caso, AG y GA dan como resultado matrices simétricas e idempotentes. Además,

$$AGA = \frac{1}{14}\begin{bmatrix} 10 & 6 & -2 \\ 6 & 5 & 3 \\ -2 & 3 & 13 \end{bmatrix}\begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 3 \\ 1 & 1 & 5 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 56 & 14 & 28 \\ 42 & 14 & 42 \\ 14 & 14 & 70 \end{bmatrix} = A \quad \text{y}$$

$$GAG = \frac{1}{20216}\begin{bmatrix} 37 & 6 & -1 \\ 6 & 2 & 6 \\ -1 & 6 & 37 \end{bmatrix}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix} = \frac{1}{532}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix} = G.$$

Así, la matriz G cumple todos los requisitos dados en (9.2).

Teorema 9.10 *Supóngase que $A = QR$ es una descomposición de la matriz A de tamaño $m \times n$, de rango $r \leq \min\{m, n\}$, de modo que Q tiene columnas ortonormales y R es triangular superior de rango r. Entonces, la inversa generalizada de A está dada por*

$$G = R^t\left(RR^t\right)^g Q^t. \tag{9.22}$$

Demostración.

Queda como ejercicio para el lector. ■

Ejemplo 9.11 *Considérese la matriz dada en el Ejemplo 9.1, obtenga la factorización QR de A y utilice el Teorema 9.10 para hallar la g-inversa.*

Solución.

La factorización QR de la matriz dada en el Ejemplo 9.1 es

$$\begin{bmatrix} 2 & -1 & 3 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{5} & -\frac{4}{5}\sqrt{5} & \frac{9}{5}\sqrt{5} \\ 0 & \frac{3}{5}\sqrt{5} & -\frac{3}{5}\sqrt{5} \end{bmatrix}.$$

Luego,

$$RR^t = \begin{bmatrix} \sqrt{5} & -\frac{4}{5}\sqrt{5} & \frac{9}{5}\sqrt{5} \\ 0 & \frac{3}{5}\sqrt{5} & -\frac{3}{5}\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ -\frac{4}{5}\sqrt{5} & \frac{3}{5}\sqrt{5} \\ \frac{9}{5}\sqrt{5} & -\frac{3}{5}\sqrt{5} \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 122 & -39 \\ -39 & 18 \end{bmatrix},$$

y por lo tanto

$$\left(RR^t\right)^g = \frac{1}{135}\begin{bmatrix} 18 & 39 \\ 39 & 122 \end{bmatrix}.$$

Por otra parte,

$$R^t\left(RR^t\right)^g = \frac{1}{135}\begin{bmatrix} \sqrt{5} & 0 \\ -\frac{4}{5}\sqrt{5} & \frac{3}{5}\sqrt{5} \\ \frac{9}{5}\sqrt{5} & -\frac{3}{5}\sqrt{5} \end{bmatrix} \begin{bmatrix} 18 & 39 \\ 39 & 122 \end{bmatrix} = \frac{\sqrt{5}}{135}\begin{bmatrix} 18 & 39 \\ 9 & 42 \\ 9 & -3 \end{bmatrix}.$$

Empleando el procedimiento dado en (9.22), se tiene que

$$G = \frac{\sqrt{5}}{135}\begin{bmatrix} 18 & 39 \\ 9 & 42 \\ 9 & -3 \end{bmatrix} \begin{bmatrix} \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} \end{bmatrix} = \frac{15}{135}\begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix},$$

la cual coincide con la obtenida en el Ejemplo 9.7.

Ejemplo 9.12 Inversa generalizada de una matriz cuadrada

Obtenga la g-inversa, usando el procedimiento dado en (9.22) *para la matriz:*

$$A = \begin{bmatrix} -1 & 1 & -2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix}.$$

Solución.

La factorización QR de la matriz dada es

$$\begin{bmatrix} -1 & 1 & -2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3}\sqrt{3} & -\frac{1}{6}\sqrt{6} & \frac{1}{2}\sqrt{2} \\ \frac{1}{3}\sqrt{3} & \frac{1}{6}\sqrt{6} & \frac{1}{2}\sqrt{2} \\ -\frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{6} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} & -\sqrt{3} & \sqrt{3} \\ 0 & 0 & \sqrt{6} \\ 0 & 0 & 0 \end{bmatrix}.$$

Luego,

$$RR^t = \begin{bmatrix} \sqrt{3} & -\sqrt{3} & \sqrt{3} \\ 0 & 0 & \sqrt{6} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ -\sqrt{3} & 0 & 0 \\ \sqrt{3} & \sqrt{6} & 0 \end{bmatrix} = \begin{bmatrix} 9 & 3\sqrt{2} & 0 \\ 3\sqrt{2} & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

y por lo tanto

$$\left(RR^t\right)^g = \frac{1}{12} \begin{bmatrix} 2 & -\sqrt{2} & 0 \\ -\sqrt{2} & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Por otra parte,

$$R^t\left(RR^t\right)^g = \frac{\sqrt{3}}{12} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 2 & -\sqrt{2} & 0 \\ -\sqrt{2} & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \frac{\sqrt{3}}{12} \begin{bmatrix} 2 & -\sqrt{2} & 0 \\ -2 & \sqrt{2} & 0 \\ 0 & 2\sqrt{2} & 0 \end{bmatrix}.$$

Mediante el procedimiento dado en (9.22), se tiene que

$$G = \frac{\sqrt{3}}{12} \begin{bmatrix} 2 & -\sqrt{2} & 0 \\ -2 & \sqrt{2} & 0 \\ 0 & 2\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{3} & -\frac{1}{3}\sqrt{3} \\ -\frac{1}{6}\sqrt{6} & \frac{1}{6}\sqrt{6} & \frac{1}{3}\sqrt{6} \\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix},$$

de manera que

$$AG = \frac{1}{12} \begin{bmatrix} -1 & 1 & -2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix},$$

$$GA = \frac{1}{12} \begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix} \begin{bmatrix} -1 & 1 & -2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix}.$$

De este modo, $AG = GA$. Además,

$$AGA = \frac{1}{12} \begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix} \begin{bmatrix} -1 & 1 & -2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} -12 & 12 & -24 \\ 12 & -12 & 24 \\ -12 & 12 & 12 \end{bmatrix} = A \quad \text{y}$$

$$GAG = \frac{1}{144} \begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix} \begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 6 & -6 & 0 \\ -6 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix} = G.$$

Así, la matriz G cumple todos los requisitos dados en (9.2).

Teorema 9.11 *Sea A una matriz real de tamaño $n \times n$ con valores singulares distintos de cero $\sigma_1, \sigma_2, \ldots, \sigma_r$ ($r = \rho(A)$). Entonces, la inversa generalizada de A está dada por*

$$G = VS^gU^t, \tag{9.23}$$

donde U y V son matrices ortogonales de tamaño $n \times n$ y S^g es la inversa generalizada de la matriz dada en (3.39).

Demostración.

Por el Teorema 3.32, la matriz A se puede expresar como $A = USV^t$ y, por consiguiente, la g-inversa es

$$G = VS^gU^t.$$

Veamos si G cumple la primera condición de la Definición 9.1:

$$\begin{aligned} AGA &= A(VS^gU^t)A = (USV^t)(VS^gU^t)(USV^t) \\ &= US(V^tV)S^g(U^tU)SV^t = USV^t = A. \end{aligned}$$

Aquí, se utilizaron los hechos de que U y V son matrices ortogonales y de que S^g es una inversa generalizada de S. Luego, G es una matriz A^{g_1}. Veamos si es A^{g_2}:

$$\begin{aligned} GAG &= (VS^gU^t)A(VS^gU^t) = (VS^gU^t)(USV^t)(VS^gU^t) \\ &= V(S^gSS^g)U^t = VS^gU^t = G. \end{aligned}$$

Ahora, observemos si G es una matriz A^{g_3}

$$AG = A(VS^gU^t) = (USV^t)(VS^gU^t) = U(SS^g)U^t. \tag{9.24}$$

Como

$$SS^g = \begin{bmatrix} I_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}, \tag{9.25}$$

la matriz AG es simétrica e idempotente, por lo tanto G también es A^{g_3}. Observemos si cumple la cuarta condición dada en (9.2),

$$GA = (VS^gU^t)A = (VS^gU^t)(USV^t) = V\big(S^gS\big)V^t. \tag{9.26}$$

Usando de nuevo el hecho dado en (9.25), se tiene

$$(GA)^t = GA.$$

Así, G es la g-inversa de A y el teorema queda demostrado. ■

Ejemplo 9.13 *Considérese la matriz dada en el Ejemplo 9.12, obtenga la g-inversa.*

Solución.

En este caso, los valores singulares de A son $\sigma_1^2 = 12$, $\sigma_2^2 = 3$ y $\sigma_3^2 = 0$. Al calcular los respectivos vectores propios normalizados de A^tA, se obtiene

$$\vec{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix}\text{-}1\\1\\\text{-}2\end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{3}}\begin{bmatrix}\text{-}1\\1\\1\end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix}.$$

Por otra parte, los respectivos vectores propios normalizados de la matriz AA^t son

$$\vec{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\\text{-}1\\0\end{bmatrix}, \qquad \vec{u}_2 = \begin{bmatrix}0\\0\\1\end{bmatrix} \qquad \text{y} \qquad \vec{u}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix}.$$

Por lo tanto, sí se establece que

$$U = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 0 & 1\\\text{-}1 & 0 & 1\\0 & \sqrt{2} & 0\end{bmatrix} \qquad \text{y} \qquad V^t = \frac{1}{\sqrt{6}}\begin{bmatrix}\text{-}1 & 1 & \text{-}2\\\text{-}\sqrt{2} & \sqrt{2} & \sqrt{2}\\\sqrt{3} & \sqrt{3} & 0\end{bmatrix}$$

y

$$S = \begin{bmatrix}\sqrt{12} & 0 & 0\\0 & \sqrt{3} & 0\\0 & 0 & 0\end{bmatrix}, \qquad \text{luego} \qquad S^g = \frac{\sqrt{3}}{6}\begin{bmatrix}1 & 0 & 0\\0 & 2 & 0\\0 & 0 & 0\end{bmatrix},$$

se obtiene

$$G = \frac{\sqrt{3}}{6}\,\frac{1}{\sqrt{6}}\,\frac{1}{\sqrt{2}}\begin{bmatrix}\text{-}1 & \text{-}\sqrt{2} & \sqrt{3}\\1 & \sqrt{2} & \sqrt{3}\\\text{-}2 & \sqrt{2} & 0\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 2 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}1 & \text{-}1 & 0\\0 & 0 & \sqrt{2}\\1 & 1 & 0\end{bmatrix}.$$

Después de realizar la multiplicación de las matrices, queda

$$G = \frac{1}{12}\begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix},$$

la cual coincide con la obtenida en el Ejemplo 9.12.

Corolario 9.11.1 *Sea A una matriz real de tamaño $m \times n$ con valores singulares distintos de cero $\sigma_1, \sigma_2, \ldots, \sigma_r$ ($r = \rho(A)$). Entonces, la g-inversa de A es la matriz definida de la siguiente forma:*

$$G = VS^gU^t, \tag{9.27}$$

donde U y V son matrices ortogonales de tamaño $m \times m$ y $n \times n$, respectivamente y S^g es la inversa generalizada de la matriz dada en (3.39).

Demostración.

Queda como ejercicio para el lector. ■

Ejemplo 9.14 *Considérese la matriz dada en el Ejemplo 9.1, obtenga la g-inversa.*

Solución.

En este caso, los valores singulares de A son $\sigma_1^2 = 27$ y $\sigma_2^2 = 1$. Al calcular los respectivos vectores propios normalizados de A^tA, se obtiene:

$$\vec{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \qquad \vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

Por otra parte, los respectivos vectores propios normalizados de la matriz AA^t son

$$\vec{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{y} \qquad \vec{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Por lo tanto, si se establece

$$U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \qquad \text{y} \qquad V^t = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 & -1 & 2 \\ -\sqrt{3} & -\sqrt{3} & 0 \\ -\sqrt{2} & \sqrt{2} & \sqrt{2} \end{bmatrix},$$

y

$$S = \begin{bmatrix} \sqrt{27} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \text{luego} \qquad S^g = \frac{1}{3\sqrt{3}}\begin{bmatrix} 1 & 0 \\ 0 & 3\sqrt{3} \\ 0 & 0 \end{bmatrix}.$$

Así se obtiene

$$G = \frac{1}{3\sqrt{3}}\,\frac{1}{\sqrt{6}}\,\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -\sqrt{3} & -\sqrt{2} \\ -1 & -\sqrt{3} & \sqrt{2} \\ 2 & 0 & \sqrt{2} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 3\sqrt{3} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Después de multiplicar las matrices, queda

$$G = \frac{1}{9}\begin{bmatrix} 5 & -4 \\ 4 & -5 \\ 1 & 1 \end{bmatrix},$$

la cual coincide con la obtenida en el Ejemplo 9.7.

El primer método para determinar inversas generalizadas fue presentado en Penrose (1955*a*, 1955*b*) y está dado en el siguiente teorema.

Teorema 9.12 Método de Penrose

Sea A una matriz real de tamaño $m \times n$ y rango $r \leq \min\{m, n\}$, particionada de la siguiente forma:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{9.28}$$

con A_{11} una submatriz no singular de tamaño $r \times r$ y $\left(A/A_{11}\right) = \mathrm{O}$. Entonces, la inversa generalizada de A es

$$G = \begin{bmatrix} A_{11}^t P_{11} A_{11}^t & A_{11}^t P_{11} A_{21}^t \\ A_{12}^t P_{11} A_{11}^t & A_{12}^t P_{11} A_{21}^t \end{bmatrix}, \tag{9.29}$$

donde $P_{11} = \left(A_{11}A_{11}^t + A_{12}A_{12}^t\right)^{-1} A_{11}\left(A_{11}^t A_{11} + A_{21}^t A_{21}\right)^{-1}$.

Demostración.

Como A_{11} es no singular, la matriz A se puede particionar como sigue:

$$\underbrace{\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}}_{A_{m\times n}} = \underbrace{\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}}_{L_{m\times r}} \underbrace{\begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}}_{U_{r\times n}}.$$

En esta partición se usa el hecho que $\big(A/A_{11}\big) = \mathrm{O}$. Luego,

$$A^g = \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^g \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^g.$$

Por el Corolario 9.8.3, se tiene que

$$\begin{aligned} \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^g &= \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^t \left\{ \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix} \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^t \right\}^{-1} \\ &= \begin{bmatrix} I_r \\ A_{12}^t\big(A_{11}^t\big)^{-1} \end{bmatrix} \left[I_r + A_{11}^{-1}A_{12}A_{12}^t\big(A_{11}^t\big)^{-1} \right]^{-1} \\ &= \begin{bmatrix} \big(I_r + A_{11}^{-1}A_{12}A_{12}^t\big(A_{11}^t\big)^{-1}\big)^{-1} \\ A_{12}^t\big(A_{11}^t\big)^{-1}\big(I_r + A_{11}^{-1}A_{12}A_{12}^t\big(A_{11}^t\big)^{-1}\big)^{-1} \end{bmatrix}. \end{aligned}$$

Nótese que

$$\begin{aligned} \left[I_r + A_{11}^{-1}A_{12}A_{12}^t\big(A_{11}^t\big)^{-1} \right]^{-1} &= \left[A_{11}^{-1}\big(A_{11} + A_{12}A_{12}^t\big(A_{11}^t\big)^{-1}\big) \right]^{-1} \\ &= \left[A_{11} + A_{12}A_{12}^t\big(A_{11}^t\big)^{-1} \right]^{-1} A_{11} \\ &= \left[\big(A_{11}A_{11}^t + A_{12}A_{12}^t\big)\big(A_{11}^t\big)^{-1} \right]^{-1} A_{11} \\ &= A_{11}^t\big(A_{11}A_{11}^t + A_{12}A_{12}^t\big)^{-1}A_{11}, \end{aligned}$$

por lo tanto,

$$\begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^g = \begin{bmatrix} A_{11}^t\big(A_{11}A_{11}^t + A_{12}A_{12}^t\big)^{-1}A_{11} \\ A_{12}^t\big(A_{11}A_{11}^t + A_{12}A_{12}^t\big)^{-1}A_{11} \end{bmatrix}.$$

Por otra parte, por el Corolario 9.8.2, se tiene que

$$\begin{aligned} \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^g &= \left\{ \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^t \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} \right\}^{-1} \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^t \\ &= \left[A_{11}^tA_{11} + A_{21}^tA_{21} \right]^{-1} \begin{bmatrix} A_{11}^t & A_{21}^t \end{bmatrix} \\ &= \begin{bmatrix} \big(A_{11}^tA_{11} + A_{21}^tA_{21}\big)^{-1}A_{11}^t & \big(A_{11}^tA_{11} + A_{21}^tA_{21}\big)^{-1}A_{21}^t \end{bmatrix}. \end{aligned}$$

Si se realizan los respectivos productos, se llega a que la A^g es

$$A^g = \begin{bmatrix} I_r & A_{11}^{-1}A_{12} \end{bmatrix}^g \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}^g = \begin{bmatrix} A_{11}^t P_{11} A_{11}^t & A_{11}^t P_{11} A_{21}^t \\ A_{12}^t P_{11} A_{11}^t & A_{12}^t P_{11} A_{21}^t \end{bmatrix}, \qquad (9.30)$$

con $P_{11} = \left(A_{11}A_{11}^t + A_{12}A_{12}^t\right)^{-1} A_{11} \left(A_{11}^t A_{11} + A_{21}^t A_{21}\right)^{-1}$ y el teorema queda demostrado. ■

Corolario 9.12.1 *Sea A una matriz cualquiera de rango r, particionada como en 9.28, donde A_{22} es una submatriz no singular de tamaño $r \times r$ y la submatriz $\left(A/A_{22}\right) = \mathrm{O}$. Entonces, la inversa generalizada de A es*

$$G = \begin{bmatrix} A_{21}^t P_{22} A_{12}^t & A_{21}^t P_{22} A_{22}^t \\ A_{22}^t P_{22} A_{12}^t & A_{22}^t P_{22} A_{22}^t \end{bmatrix}, \qquad (9.31)$$

donde $P_{22} = \left(A_{22}A_{22}^t + A_{21}A_{21}^t\right)^{-1} A_{22} \left(A_{22}^t A_{22} + A_{12}^t A_{12}\right)^{-1}$.

Demostración.

Queda como ejercicio para el lector. ■

Antes de dar ejemplos, se proporciona el siguiente procedimiento que recoge el método de Penrose para determinar la inversa generalizada:

Procedimiento para determinar la inversa generalizada

Realice una partición de la matriz A como sigue:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

de tal manera que una de las submatrices A_{11} o A_{22} sea cuadrada y tenga igual rango que la matriz A.

I. Si A_{11} es la submatriz no singular:

a) Verifique que $\left(A/A_{11}\right) = \mathrm{O}$.

b) Obtenga:

$$P_{11} = \left(A_{11}A_{11}^t + A_{12}A_{12}^t\right)^{-1} A_{11} \left(A_{11}^t A_{11} + A_{21}^t A_{21}\right)^{-1}.$$

c) Forme la matriz:

$$\begin{bmatrix} A_{11}^t P_{11} A_{11}^t & A_{11}^t P_{11} A_{21}^t \\ A_{12}^t P_{11} A_{11}^t & A_{12}^t P_{11} A_{21}^t \end{bmatrix},$$

este resultado es la inversa generalizada de A.

II. Si A_{22} es la submatriz no singular:

a) Verifique que $\big(A/A_{22}\big) = \mathrm{O}$.

b) Obtenga:

$$P_{22} = \big(A_{22}A_{22}^t + A_{21}A_{21}^t\big)^{-1} A_{22}\big(A_{22}^t A_{22} + A_{12}^t A_{12}\big)^{-1}.$$

c) Forme la matriz:

$$\begin{bmatrix} A_{21}^t P_{22} A_{12}^t & A_{21}^t P_{22} A_{22}^t \\ A_{22}^t P_{22} A_{12}^t & A_{22}^t P_{22} A_{22}^t, \end{bmatrix},$$

este resultado es la inversa generalizada de A.

Ejemplo 9.15 *Calcule, utilizando el método descrito anteriormente, la inversa generalizada de la matriz dada en el Ejemplo 9.4.*

Solución.

Para la partición $(2+1)\times(2+1)$, de la matriz dada se tiene

$$A = \left[\begin{array}{ccc:c} 4 & 1 & \vdots & 2 \\ 3 & 1 & \vdots & 3 \\ \hdots & \hdots & . & \hdots \\ 1 & 1 & \vdots & 5 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Luego, A_{11} es no singular ya que su determinante es 1. El lector puede verificar que $\big(A/A_{11}\big) = \mathrm{O}$. Además,

$$A_{11}A_{11}^t + A_{12}A_{12}^t = \begin{bmatrix} 17 & 13 \\ 13 & 10 \end{bmatrix} + \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} = \begin{bmatrix} 21 & 19 \\ 19 & 19 \end{bmatrix}$$

y

$$A_{11}^t A_{11} + A_{21}^t A_{21} = \begin{bmatrix} 25 & 7 \\ 7 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 26 & 8 \\ 8 & 3 \end{bmatrix}.$$

Por lo tanto, P_{11} se obtiene como

$$P_{11} = \begin{bmatrix} 21 & 19 \\ 19 & 19 \end{bmatrix}^{-1} \begin{bmatrix} 4 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 26 & 8 \\ 8 & 3 \end{bmatrix}^{-1} = \frac{1}{532} \begin{bmatrix} 57 & -152 \\ -55 & 156 \end{bmatrix}.$$

El lector puede realizar los otros productos y llegar a que

$$G = \frac{1}{532}\begin{bmatrix} 112 & 49 & -77 \\ 12 & 10 & 6 \\ -40 & 11 & 113 \end{bmatrix},$$

la cual coincide con la obtenida en el Ejemplo 9.10.

Ejemplo 9.16 *Considérese la matriz dada en el Ejemplo 9.12. Obtenga la g-inversa mediante el método descrito anteriormente.*

Solución.

Realizando una partición $(1+2)\times(1+2)$ a la matriz dada, se tiene

$$A = \left[\begin{array}{c:cc} -1 & 1 & -2 \\ \hdashline \cdots & \cdots & \cdots \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Luego, A_{22} es no singular, ya que su determinante es -3. El lector puede verificar que $\left(A/A_{22}\right) = \mathrm{O}$. También,

$$A_{22}A_{22}^t + A_{21}A_{21}^t = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix}$$

y

$$A_{22}^tA_{22} + A_{12}^tA_{12} = \begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 3 & -3 \\ -3 & 9 \end{bmatrix}.$$

Por lo tanto, P_{22} se obtiene como

$$P_{22} = \begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix}^{-1}\begin{bmatrix} -1 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & -3 \\ -3 & 9 \end{bmatrix}^{-1} = \frac{1}{36}\begin{bmatrix} -1 & 1 \\ 8 & 4 \end{bmatrix}.$$

El lector puede realizar los otros productos y llegar a que

$$G = \frac{1}{12}\begin{bmatrix} -1 & 1 & -4 \\ 1 & -1 & 4 \\ -2 & 2 & 4 \end{bmatrix}.$$

Nótese que esta matriz coincide con la obtenida en el Ejemplo 9.12.

9.4 Vectores y valores propios

Si A es una matriz no singular, por el Teorema 2.10 se sabe que los valores propios de A^{-1} son los recíprocos de los valores propios de A y los correspondientes vectores propios son los mismos. En esta sección se muestran las relaciones entre los valores y vectores propios de una matriz cuadrada y los asociados a la g-inversa.

Teorema 9.13 *Sea G la g-inversa de A y λ un valor propio distinto de cero de A con vector propio correspondiente $\vec{v} \neq \vec{0}$. Entonces, una condición suficiente para que $G\vec{v} = \lambda^{-1}\vec{v}$ es que $AG = GA$.*

Demostración.

Si $A\vec{v} = \lambda\vec{v}$, entonces, premultiplicando por AG, se obtiene que

$$\begin{aligned} \underbrace{AGA}\vec{v} &= \lambda AG\vec{v} \\ A\ \ \vec{v} &= \lambda AG\vec{v} \\ \lambda\vec{v} &= \lambda AG\vec{v}. \end{aligned}$$

Como $\lambda \neq 0$, entonces

$$\begin{aligned} \vec{v} &= AG\vec{v} && \text{si } AG = GA \\ &= GA\vec{v} = \lambda G\vec{v}. \end{aligned}$$

Es decir, $G\vec{v} = \lambda^{-1}\vec{v}$. ■

Ejemplo 9.17 *Determine los vectores y valores propios de la matriz g_2-inversa, obtenida en el Ejemplo 9.4.*

Solución.

En el Ejemplo 9.4, se obtuvo que

$$G = \frac{1}{441}\begin{bmatrix} 106 & 9 & -52 \\ 57 & 9 & -3 \\ -41 & 9 & 95 \end{bmatrix}.$$

El polinomio característico de G es

$$p_G(\lambda) = -\frac{1}{21}\lambda\big(7\lambda - 1\big)\big(3\lambda - 1\big),$$

luego, los valores propios de G son $\lambda_1 = 0$, $\lambda_2 = \frac{1}{7}$ y $\lambda_3 = \frac{1}{3}$, y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 1 \\ -6 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} -7 \\ -3 \\ 5 \end{bmatrix}$, respectivamente.

Teorema 9.14 *Si A es simétrica, los valores propios no nulos de A y A^{g_3} son recíprocos.*

Demostración.

Si $A\vec{v} = \lambda\vec{v}$, entonces, premultiplicando por AA^{g_3}, da

$$\begin{aligned} \underbrace{AA^{g_3}A}\vec{v} &= \lambda AA^{g_3}\vec{v} \\ A\ \ \vec{v} &= \lambda\big(AA^{g_3}\big)^t\vec{v} \\ \lambda\vec{v} &= \lambda\big(A^{g_3}\big)^t A\vec{v}. \end{aligned}$$

Como $\lambda \neq 0$, entonces

$$\big(A^{g_3}\big)^t\vec{v} = \lambda^{-1}\vec{v}.$$

El resultado se sigue, puesto que una matriz y su transpuesta tienen los mismos valores propios. ■

Ejercicios 9.1

1. En los siguientes problemas, determine la inversa generalizada con los métodos descritos en esta sección:

$a)$ $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. $b)$ $\begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & -2 \end{bmatrix}$. $c)$ $\begin{bmatrix} 1 & 0 & 3 \\ 2 & -1 & 6 \end{bmatrix}$.

$d)$ $\begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & -1 & -3 \end{bmatrix}$. $e)$ $\begin{bmatrix} -5 & 2 & 1 \\ 2 & -2 & 2 \\ 1 & 2 & -5 \end{bmatrix}$. $f)$ $\begin{bmatrix} 3 & -1 & 4 & 1 \\ -1 & 1 & -5 & -1 \\ 4 & -5 & 2 & 0 \end{bmatrix}$.

2. Encuentre una inversa generalizada para cada una de las siguientes matrices:

(a) PAQ cuando P y Q son no singulares

(b) GA cuando G es una inversa generalizada de A

(c) kA cuando k es un escalar

(d) PAP^t cuando P es ortogonal y A es idempotente

3. Sean A y X matrices simétricas tal que $AX = 0$, si X es idempotente y $A+X$ es no singular, pruebe que $\left(A+X\right)^{-1}$ es una inversa generalizada para A y X.

4. Para X particionada como $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ con X_1 de rango completo columna, pruebe que $X(X^tX)^gX^t = X_1(X_1^tX_1)^gX_1^t$.

9.5 Solución de sistemas de ecuaciones lineales

Entre las múltiples aplicaciones que tiene el concepto de inversa generalizada cabe señalar el papel que desempeña en el análisis y la "solución" de sistemas lineales consistentes e inconsistentes. En el caso de sistemas consistentes, como las soluciones existen, permite caracterizarlas, y para sistemas inconsistentes, a pesar de que por definición carece de soluciones, proporciona soluciones aproximadas. En esta sección se analiza con ayuda de la g-inversa de la matriz A cuando (9.1) es consistente y cómo son sus soluciones.

Teorema 9.15 *El sistema de ecuaciones lineales dado en* (9.1) *es consistente si y solo si se verifica que*

$$AA^{g_1}\vec{b} = \vec{b}. \tag{9.32}$$

Demostración.
Si para A^{g_1}, se cumple que

$$AA^{g_1}\vec{b} = \vec{b},$$

entonces el sistema (9.1) es consistente, puesto que al menos $\vec{x}' = A^{g_1}\vec{b}$ es solución.

Recíprocamente, la condición es necesaria pues si el sistema es consistente, existe $\vec{x}_0 \in \mathbb{R}^n$ tal que

$$A\vec{x}_0 = \vec{b}.$$

Ahora bien, como A^{g_1} siempre existe, premultiplicando la expresión anterior por AA^{g_1}, se obtiene

$$\begin{aligned} \underbrace{AA^{g_1}A}\vec{x}_0 &= AA^{g_1}\vec{b} \\ \underbrace{A \quad \vec{x}_0} &= AA^{g_1}\vec{b} \\ \vec{b} \quad &= AA^{g_1}\vec{b}, \end{aligned}$$

lo cual prueba el teorema. ■

Teorema 9.16 *Dado un sistema consistente:*

$$A\vec{x} = \vec{b},$$

con A de tamaño $m \times n$, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, se verifica que:

i) Para todo $\vec{d} \in \mathbb{R}^n$,

$$\vec{x}_0 = A^{g_1}\vec{b} + \big(I_n - A^{g_1}A\big)\vec{d} \tag{9.33}$$

es solución del sistema.

ii) Si $\vec{x}'$ es una solución cualquiera del sistema, existe $\vec{d} \in \mathbb{R}^n$ tal que $\vec{x}'$ puede expresarse en la forma dada en (9.33).

Demostración.

i) Sea $\vec{x}_0$ una solución del sistema, es decir

$$A\vec{x}_0 = \vec{b}.$$

Entonces, por el Teorema 9.15 y la definición de A^{g_1}, se tiene que

$$A\vec{x}_0 = AA^{g_1}\vec{b} + A\big(I_n - A^{g_1}A\big)\vec{d} = \vec{b} + A\vec{d} - A\vec{d}.$$

ii) Si $\vec{x}'$ es una solución cualquiera del sistema, se verifica que

$$\vec{b} - A\vec{x}' = \vec{0}.$$

Si se premultiplica por A^{g_1} y se suma $\vec{x}'$ a ambos lados, se tiene que

$$\begin{aligned} A^{g_1}\vec{b} - A^{g_1}A\vec{x}' + \vec{x}' &= \vec{x}' \\ A^{g_1}\vec{b} + \big(I_n - A^{g_1}A\big)\vec{x}' &= \vec{x}'. \end{aligned}$$

Luego, tomando en el lado izquierdo $\vec{d} = \vec{x}'$, se obtiene lo que se deseaba. ■

Teorema 9.17 *Dado el sistema consistente:*

$$A\vec{x} = \vec{b},$$

con A de tamaño $m \times n$, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, se verifica que existe solución única $\vec{x}'$ si y solo si $A^g A = I_n$, siendo A^g la g-inversa de la matriz A.

Demostración.

Queda como ejercicio para el lector. ■

Ejemplo 9.18 *Determine una solución del sistema de ecuaciones:*

$$\begin{aligned} x_1 + 6x_2 + 11x_3 &= 0 \\ 6x_1 + 46x_2 + 86x_3 &= 20 \\ 11x_1 + 86x_2 + 161x_3 &= 40. \end{aligned} \tag{9.34}$$

Solución.

Si se reescribe (9.34), se llega a

$$\begin{bmatrix} 1 & 6 & 11 \\ 6 & 46 & 86 \\ 11 & 86 & 161 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 20 \\ 40 \end{bmatrix}. \tag{9.35}$$

El sistema dado en (9.34) se puede resolver usando inversa generalizada y usando cualquiera de los métodos descritos en este capítulo. De esta manera, se tiene que

$$A^g = \frac{1}{180} \begin{bmatrix} 517 & 190 & -137 \\ 190 & 70 & -50 \\ -137 & -50 & 37 \end{bmatrix}.$$

Veamos si el sistema de ecuaciones (9.34) es consistente determinado

$$\begin{aligned} A^g A &= \frac{1}{180} \begin{bmatrix} 517 & 190 & -137 \\ 190 & 70 & -50 \\ -137 & -50 & 37 \end{bmatrix} \begin{bmatrix} 1 & 6 & 11 \\ 6 & 46 & 86 \\ 11 & 86 & 161 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} \neq I_3. \end{aligned}$$

Por el Teorema 9.17, el sistema es consistente indeterminado, luego una solución de (9.34) es

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= \frac{1}{180} \begin{bmatrix} 517 & 190 & -137 \\ 190 & 70 & -50 \\ -137 & -50 & 37 \end{bmatrix} \begin{bmatrix} 0 \\ 20 \\ 40 \end{bmatrix} \\ &= \frac{1}{9} \begin{bmatrix} 517 & 190 & -137 \\ 190 & 70 & -50 \\ -137 & -50 & 37 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = -\frac{1}{3} \begin{bmatrix} 28 \\ 10 \\ -8 \end{bmatrix}. \end{aligned} \tag{9.36}$$

Teniendo en cuenta el Teorema 9.16, para todo $\vec{d} = (\alpha, \beta, \gamma) \in \mathbb{R}^3$, el vector

$$\vec{x}_0 = A^g \vec{b} + \big(I_n - A^g A\big)\vec{d} = \frac{1}{6} \begin{bmatrix} -56 + \alpha - 2\beta + \gamma \\ -20 - 2\alpha + 4\beta - 2\gamma \\ 16 + \alpha - 2\beta + \gamma \end{bmatrix},$$

es una solución del sistema, como el lector puede comprobar fácilmente sustituyendo $\vec{x}_0$ en (9.34).

Teorema 9.18 *Dado el sistema de ecuaciones lineales:*

$$A\vec{x} = \vec{b},$$

en donde A es una matriz de tamaño $m \times n$ con rango r, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$. Considérese la función residuo:

$$\boldsymbol{r}(\vec{x}) = A\vec{x} - \vec{b}$$

para cualquier $\vec{x}$. Entonces, $\vec{x}' = A^g\vec{b}$ es una solución aproximada mínimo cuadrática (LS por su sigla en inglés) del sistema si y solo si mínimiza a $\|\boldsymbol{r}(\vec{x})\|$.

Demostración.

Si se reescribe la norma euclídea de $\boldsymbol{r}(\vec{x})$, se obtiene que

$$\|\boldsymbol{r}(\vec{x})\|^2 = \underbrace{\boldsymbol{r}(\vec{x})^t\boldsymbol{r}(\vec{x})}_{F(\vec{x})} = (A\vec{x} - \vec{b})^t(A\vec{x} - \vec{b}) = (\vec{x}^tA^t - \vec{b}^t)(A\vec{x} - \vec{b})$$
$$= \vec{x}^tA^tA\vec{x} - \vec{b}^tA\vec{x} - \vec{x}^tA^t\vec{b} + \vec{b}^t\vec{b}.$$

Para determinar la $\vec{x}$ que minimiza esta suma de cuadrados de los residuos, se calculan las derivadas parciales[2] de $F(\vec{x})$ con respecto a $\vec{x}$

$$\frac{\partial F(\vec{x})}{\partial \vec{x}} = 2A^tA\vec{x} - 2A^t\vec{b} = \vec{0}. \tag{9.37}$$

Al despejar $\vec{x}$, se obtiene un mínimo global de $F(\vec{x})$ pues

$$\frac{\partial^2 F(\vec{x})}{\partial \vec{x}^2} = 2A^tA,$$

la cual es una matriz definida positiva si A es de rango completo, o semidefinida positiva en caso contrario, y por ello, en ambas situaciones, $F(\vec{x})$ es una función convexa.

Luego, si se sustituye $\vec{x}'$ en (9.37), se tiene

$$A^tAA^g\vec{b} - A^t\vec{b} = \vec{0} \qquad \text{o} \qquad (A^tAA^g - A^t)\vec{b} = \vec{0},$$

[2] Si el lector desea consultar técnicas de derivación matricial, puede ver Barbolla & Sanz (1998, cap. 5).

pero como $\vec{b}$ es cualquier vector de $\mathbb{R}^m$, esto equivale a la condición

$$A^t A A^g = A^t;$$

esto se sigue inmediatamente del Teorema 9.3 parte d). ■

Definición 9.2 *Dado el sistema lineal* $A\vec{x} = \vec{b}$ *y la función residuo* $\boldsymbol{r}(\vec{x}) = A\vec{x} - \vec{b}$, *se dice que:*

i) $\vec{x}'$ *es una solución aproximada mínimo cuadrática (LS por su sigla en inglés) del sistema si y solo si para todo* $\vec{x} \in \mathbb{R}^n$, *se verifica que*

$$\boldsymbol{r}(\vec{x}')^t \boldsymbol{r}(\vec{x}') \leq \boldsymbol{r}(\vec{x})^t \boldsymbol{r}(\vec{x}).$$

ii) *Una solución aproximada mínimo cuadrática* $\vec{x}'$ *es de norma mínima (MNLS por su sigla en inglés) si y solo si para todo* $\vec{x} \in \mathbb{R}^n$, *se cumple que*

$$\boldsymbol{r}(\vec{x}')^t \boldsymbol{r}(\vec{x}') = \boldsymbol{r}(\vec{x})^t \boldsymbol{r}(\vec{x}).$$

En el siguiente resultado, se recogen dos características de las soluciones mínimo cuadráticas LS para sistemas inconsistentes a partir de la g_3-inversa de la matriz de coeficientes del sistema.

Teorema 9.19 *Dado el sistema de ecuaciones lineales inconsistente:*

$$A\vec{x} = \vec{b},$$

con A *de tamaño* $m \times n$, $\vec{x} \in \mathbb{R}^n$ *y* $\vec{b} \in \mathbb{R}^m$, *se verifica que*

(i) $\vec{x}' = G\vec{b}$ *es una solución LS del sistema si y solo si* G *es una* A^{g_3}.

(ii) *La solución* $\vec{x}_0$ *dada en* (9.33) *es una solución LS del sistema si y solo si* $\vec{x}'$ *es solución del sistema consistente:*

$$A\vec{x} = AA^{g_3}\vec{b}.$$

Demostración.

(i) Por el Teorema 9.18, una solución LS del sistema es de la forma

$$\vec{x} = A^g\vec{b}.$$

Falta entonces comprobar qué condiciones de las dadas en (9.2) debe cumplir G si $\vec{x}' = G\vec{b}$ es solución LS del sistema y, por tanto, solución de (9.37). Así, sustituyendo $\vec{x}'$ en (9.37), se tiene

$$A^tAG\vec{b} - A^t\vec{b} = \vec{0},$$

y como $\vec{b}$ es cualquier vector de $\mathbb{R}^m$, esto equivale a la condición

$$A^tAG = A^t,$$

que se verifica si y solo si G es una g_3-inversa de A, ya que

$$AG \text{ es simétrica} \qquad \text{y} \qquad A^tG^tA^t = A^t.$$

(ii) Si el sistema

$$A\vec{x} = AA^{g_3}\vec{b}$$

es consistente, entonces por el Teorema 9.16, sus soluciones son de la forma

$$\vec{x}' = A^{g_1}\big(AA^{g_3}\vec{b}\big) + \big(I_n - A^{g_1}A\big)\vec{d}$$

para cualquier $\vec{d} \in \mathbb{R}^m$ o también

$$\vec{x}' = A^{g_3}\big(AA^{g_3}\vec{b}\big) + \big(I_n - A^{g_3}A\big)\vec{d} = A^{g_3}\vec{b} + \big(I_n - A^{g_3}A\big)\vec{d},$$

como cualquier g_3-inversa de A es a su vez g_1-inversa.

La demostración concluye si se muestra que $\vec{x}'$ es una solución LS del sistema $A\vec{x} = \vec{b}$. Para ello, razonando como en (i), $\vec{x}'$ debe ser solución de (9.37), lo cual es válido pues de acuerdo con la definición de la g_3-inversa,

$$A^tA\big[A^{g_3}\vec{b} + \big(I_n - A^{g_3}A\big)\vec{d}\big] - A^t\vec{b} = A^tAA^{g_3}\vec{b} - A^t\vec{b},$$

y como $AA^{g_3} = \big(A^{g_3}\big)^tA^t$, dada la simetría de AA^{g_3}, se obtiene finalmente que

$$A^t\big(A^{g_3}\big)^tA^t\vec{b} - A^t\vec{b} = \vec{0}. \quad \blacksquare$$

Teorema 9.20 *Dado el sistema inconsistente:*

$$A\vec{x} = \vec{b},$$

con A de tamaño $m \times n$, $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$, su solución $MNLS$ es única y está dada por

$$\vec{x}' = A^g\vec{b}. \tag{9.38}$$

Demostración.

Si A^g es la g-inversa de A, también es g_3-inversa de A y, por ello, en virtud del teorema anterior, está garantizado que

$$\vec{x}' = A^g\vec{b}$$

es solución LS del sistema.

Bastará, por tanto, probar que $\vec{x}'$ es única y de mínima norma.

- $\vec{x}'$ es $MNLS$.

  En efecto, por ser $\vec{x}'$ LS, es solución del sistema

  $$A\vec{x} = AA^g\vec{b}$$

  o equivalentemente

  $$A^gA\vec{x} = A^gAA^g\vec{b} = A^g\vec{b}. \tag{9.39}$$

  Por lo tanto, cualquier solución LS es de la forma

  $$\vec{x} = A^g\vec{b} + \big(I_n - A^gA\big)\vec{x},$$

  y el cuadrado de su norma euclídea es

  $$\vec{x}^t\vec{x} = \vec{b}^t\big(A^g\big)^tA^g\vec{b} + \big(\vec{x} - A^gA\vec{x}\big)^t\big(\vec{x} - A^gA\vec{x}\big).$$

  Si se sustituye (9.39) en esta expresión, resulta

  $$\begin{aligned}\vec{x}^t\vec{x} &= \vec{b}^t\big(A^g\big)^tA^g\vec{b} + \big(\vec{x} - A^g\vec{b}\big)^t\big(\vec{x} - A^g\vec{b}\big)\\ &= \big\|A^g\vec{b}\big\|^2 + \big\|\vec{x} - A^g\vec{b}\big\|^2,\end{aligned}$$

  de donde

  $$\vec{x}^t\vec{x} \geq \big(\vec{x}'\big)^t\big(\vec{x}'\big) = \big\|A^g\vec{b}\big\|^2$$

  cuando $\vec{x}' = A^g\vec{b}$.

- $\vec{x}'$ es única.

  Supóngase que $\vec{x}_0 \in \mathbb{R}^n$ es también solución $MNLS$ del sistema. Entonces, $\vec{x}_0$ cumple que

$$\left(\vec{x}_0\right)^t\left(\vec{x}_0\right) = \left\|A^g\vec{b}\right\|^2 + \left\|\vec{x}_0 - A^g\vec{b}\right\|^2 = \left(\vec{x}'\right)^t\left(\vec{x}'\right) = \left\|A^g\vec{b}\right\|^2.$$

  Por lo tanto,

$$\left\|\vec{x}_0 - A^g\vec{b}\right\|^2 = 0.$$

  Es decir,

$$\vec{x}_0 - A^g\vec{b} = \vec{0} \qquad \Leftrightarrow \qquad \vec{x}_0 = A^g\vec{b}.$$

  En consecuencia,

$$\vec{x}_0 = \vec{x}',$$

y esto es lo que se quería demostrar. ■

Ejercicios 9.2

1. Encuentre una solución para cada uno de los sistemas de ecuaciones lineales dados a continuación:

$a)$
$$\begin{aligned} 3x_1 - 2x_2 - x_3 &= 1 \\ -x_1 + 2x_2 + 2x_3 &= 2 \\ x_1 + 2x_2 + 3x_3 &= 4. \end{aligned}$$

$b)$
$$\begin{aligned} x_1 - x_2 - x_3 &= 2 \\ 2x_1 + x_2 + 2x_3 &= 4 \\ x_1 - 4x_2 - 5x_3 &= 2. \end{aligned}$$

$c)$
$$\begin{aligned} x_1 - 2x_2 + 3x_3 + x_4 &= 2 \\ x_1 \quad - x_3 + x_4 &= 5 \\ 3x_1 - 2x_2 + 2x_3 \quad &= 2. \end{aligned}$$

$d)$
$$\begin{aligned} x_1 - 2x_2 + x_3 + x_4 &= 2 \\ 3x_1 \quad + 2x_3 - 2x_4 &= -8 \\ 4x_2 - x_3 - x_4 &= 1 \\ 5x_1 \quad + 3x_3 - x_4 &= -3. \end{aligned}$$

2. Sea X una matriz de tamaño $m \times n$ $(m > n)$ y rango $r < n$. Sea G una inversa generalizada de X^tX, defina:

$$\vec{b} = GX^t\vec{Y}, \qquad s^2 = (\vec{Y} - X\vec{b})^t(\vec{Y} - X\vec{b}),$$

$$\vec{b}_0 = \vec{b} - GQ\big(Q^tGQ\big)^{-1}\big(Q^t\vec{b} - \vec{m}\big), \qquad \text{con} \quad Q = \big(GX^tXG^t\big)^tX.$$

Pruebe que:

(a) $s^2 = \vec{Y}^t\vec{Y} - \vec{b}^tX^t\vec{Y}$.

(b) $Q^t\vec{b}_0 = \vec{m}$.

(c) $(\vec{Y} - X\vec{b}_0)^t(\vec{Y} - X\vec{b}_0) = s^2 + (Q^t\vec{b} - \vec{m})^t(Q^tGQ)^{-1}(Q^t\vec{b} - \vec{m})$.

Capítulo 10

Aplicaciones

En este capítulo se recopilan algunos desarrollos teóricos de la estadística. El lector que esté interesado en profundizarlos puede consultar textos de modelos lineales o estadística multivariada (o cursar las asignaturas correspondientes). El propósito de este capítulo es ilustrar la utilidad de los conceptos tratados en este escrito, y por eso se omiten tanto conceptos básicos del área de la estadística, como aquellos temas avanzados que el lector aprenderá posteriormente.

10.1 Matrices estocásticas

Las matrices estocásticas corresponden a un tipo especial de matrices definidas positivas, y se usan con frecuencia en el estudio de fenómenos aleatorios, en teoría de la probabilidad y estadística.

Definición 10.1 *Una matriz* $A = [a_{ij}]$ *de tamaño* $n \times n$ *se dice que es estocástica por filas (columnas) si todos sus elementos son números reales no negativos y la suma de los elementos de cada una de sus filas (columnas) es igual a* 1. *Es decir:*

$$0 \leq a_{ij} \leq 1 \qquad\qquad i, j = 1, 2, \ldots, n$$

y además:

$$A\mathbf{1} = \sum_{j=1}^{n} a_{ij} = 1 \qquad \textit{si } A \textit{ es estocástica por filas}$$

$$\mathbf{1}^t A = \sum_{i=1}^{n} a_{ij} = 1 \qquad \textit{si } A \textit{ es estocástica por columnas,}$$

donde $\mathbf{1}$ *es el vector columna de unos de tamaño* $n \times 1$. *Se dice que* A *es doblemente estocástica cuando es estocástica tanto por filas como por columnas.*

Teorema 10.1 *Si* A *y* B *son estocásticas (doblemente estocásticas), se verifica:*

i) AB *es estocástica (doblemente estocástica).*

ii) Para todo $k \in \mathbb{N}$, A^k *es estocástica (doblemente estocástica).*

iii) Cuando A *es doblemente estocástica, entonces* A^t *también lo es.*

Teorema 10.2 *Si* A *es una matriz estocástica por filas (columnas), entonces* $\lambda = 1$ *es uno de sus valores propios.*

Demostración.

Sea A una matriz de tamaño $n \times n$ tal que A es una matriz estocástica por columnas. Basta probar que

$$\det\big(A - I\big) = 0.$$

Para ello, veamos que las filas de la matriz $A - I$ no son linealmente independientes. Si $B = A - I$, consideremos la suma vectorial de las filas de la matriz B:

$$\vec{B}_1^t + \vec{B}_2^t + \ldots \vec{B}_n^t = \begin{bmatrix} a_{11} - 1 \\ a_{12} \\ \vdots \\ a_{1n} \end{bmatrix} + \begin{bmatrix} a_{21} \\ a_{22} - 1 \\ \vdots \\ a_{2n} \end{bmatrix} + \ldots + \begin{bmatrix} a_{n1} \\ a_{n2} \\ \vdots \\ a_{nn} - 1 \end{bmatrix}.$$

Es decir,

$$\vec{B}_1^t + \vec{B}_2^t + \ldots \vec{B}_n^t = \begin{bmatrix} a_{11} + a_{21} + \ldots + a_{n1} - 1 \\ a_{12} + a_{22} + \ldots + a_{n2} - 1 \\ \vdots \\ a_{1n} + a_{2n} + \ldots + a_{nn} - 1 \end{bmatrix}. \tag{10.1}$$

Como A es una matriz estocástica por columnas, las entradas de cada columna de A suman uno, y por lo tanto

$$\sum_{i=1}^{n} a_{ij} = 1 \qquad \forall j,$$

luego, (10.1) se transforma en $\vec{B}_1^t + \vec{B}_2^t + \ldots + \vec{B}_n^t = \vec{0}$, es decir, se encontró una combinación lineal no trivial de las filas de $B = A - I$ que producen el vector cero de $\mathbb{R}^n$. Por lo tanto, las filas de $A - I$ son linealmente dependientes, luego la matriz $A - I$ es singular, es decir, $\det\big(A - I\big) = 0$ y entonces, $\lambda = 1$ es un valor propio de A. ■

Definición 10.2 Matriz regular

Una matriz estocástica A se dice regular si todos los elementos de al menos una de sus potencias A^k (k entero positivo) son estrictamente positivos (mayores que cero).

Definición 10.3 Cadena de Markov

Una cadena de Markov, o proceso de Markov, es un proceso en el cual la probabilidad de que el sistema esté en un estado particular en un periodo dado de observación depende solamente de su estado en el periodo de observación inmediatamente anterior.

Definición 10.4 Probabilidad de transición

Se define la probabilidad de transición p_{ij} $(i, j = 1, 2, ..., n)$ como la probabilidad de que el sistema pase del estado j al estado i en la siguiente observación.

Definición 10.5 Matriz de transición

A cada cadena de Markov se le puede asignar una única matriz de transición P, cuyos elementos son las probabilidades p_{ij}. Esta matriz es cuadrada y su dimensión depende del número posible de estados, la matriz P resulta ser estocástica.

Definición 10.6 Vector de probabilidad

Un vector de probabilidad es un vector columna con entradas no negativas en el que la suma de sus elementos es igual a la unidad. Se dice que los vectores de probabilidad $X^{(n)}$, para $n = 0, 1, \ldots,$ son los vectores de estado de un proceso de Markov si la componente de orden i, $p_i^{(n)}$ de $X^{(n)}$ es la probabilidad de que el sistema esté en el estado i cuando se hace la observación n.

Teorema 10.3 *Si P es la matriz de transición de un proceso de Markov y $X^{(n)}$ es el vector columna de la observación n, se tendrá que:*

$$X^{(n)} = \begin{cases} PX^{(n-1)} & \text{si } P \text{ es estocástica por columnas} \\ P^t X^{(n-1)} & \text{si } P \text{ es estocástica por filas} \end{cases} \tag{10.2}$$

La ecuación (10.2) implica

$$\begin{aligned} X^{(1)} &= PX^{(0)} \\ X^{(2)} &= PX^{(1)} = P\big(PX^{(0)}\big) = P^2X^{(0)} \\ X^{(3)} &= PX^{(2)} = P\big(P^2X^{(0)}\big) = P^3X^{(0)} \end{aligned}$$

y, en general,

$$X^{(n)} = P^n X^{(0)}. \tag{10.3}$$

Así, la matriz de transición y el vector de estados inicial $X^{(0)}$ determinan completamente los demás vectores de estado.

Definición 10.7 *Un proceso de Markov es regular si su matriz de transición es una matriz estocástica regular.*

Teorema 10.4 *Si P es una matriz de transición regular de tamaño $n \times n$, entonces cuando $n \to \infty$, P^n tiende a una matriz R de tamaño $n \times n$ de la forma:*

$$R = [\vec{v} \;\; \vec{v} \;\; \ldots \;\; \vec{v}],$$

donde $\vec{v}$ es un vector de probabilidad de tamaño $n \times 1$, con todos sus elementos mayores que cero.

Demostración.
El lector puede consultarla en Kemeny & Snell (1976). ■

Teorema 10.5 *Si P es una matriz de transición regular de tamaño $n \times n$ y R y $\vec{v}$ son como en el Teorema 10.4, entonces:*

(i) Para cualquier vector $X^{(0)}$ de probabilidad inicial, $P^n X^{(0)}$ tiende a $\vec{v}$ cuando aumenta n, esto es

$$\lim_{n \to \infty} \left(P^n X^{(0)}\right) = \vec{v}.$$

Es decir, todo proceso regular de Markov tiene un vector estacionario $\vec{v}$.

(ii) El vector estacionario $\vec{v}$ es el único vector de probabilidad que satisface la ecuación:

$$P\vec{v} = \vec{v}, \qquad o \qquad \left(P - I\right)\vec{v} = \vec{0}.$$

Luego, $\vec{v}$ es un vector propio de P asociado al valor propio $\lambda = 1$.

Ejemplo 10.1 *Una empresa de investigación de mercado estudia un grupo de consumidores de café, los cuales compran una lata del grano cada semana. Las personas que actualmente toman la marca A, la comprarán de nuevo la próxima semana con una probabilidad de* 0.50*, cambiarán a la marca B con una probabilidad de* 0.25 *y comprarán la marca D con una probabilidad de* 0.25*. De las personas que ahora consumen la marca B, preferirán la próxima semana la marca A, B o D con probabilidades de* $0.60, 0.30, 0.10$*, respectivamente. Ahora, de las personas que en la actualidad compran la marca D adquirirán la próxima semana la marca A, B o D con probabilidades de* $0.30, 0.40, 0.30$*. Suponga que al iniciar el estudio, la marca A tiene el* $20\,\%$ *del mercado, la marca B tiene el* $20\,\%$ *y la otra marca el* $60\,\%$*. ¿A largo plazo cuál será el porcentaje del mercado que tendrán las marcas A, B y D?*

Solución.

Si se aborda el problema por medio de las cadenas de Markov.

$$P = \begin{array}{c} \begin{array}{ccc} A & B & D \end{array} \\ \left[\begin{array}{ccc} 0.50 & 0.60 & 0.40 \\ 0.25 & 0.30 & 0.30 \\ 0.25 & 0.10 & 0.30 \end{array}\right] \end{array} \begin{array}{c} A \\ B \\ D \end{array}$$

Como P es estocástica por columnas, al calcular los valores propios de P, se obtienen $\lambda_1 = 1$, $\lambda_2 = \frac{1}{20} - \frac{\sqrt{3}}{20}i$ y $\lambda_3 = \frac{1}{20} + \frac{\sqrt{3}}{20}i$, y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 46 \\ 25 \\ 20 \end{bmatrix}$ y $\vec{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + i \begin{bmatrix} \sqrt{3} \\ 0 \\ -\sqrt{3} \end{bmatrix}$, respectivamente.

Esto implica que la matriz P no es diagonalizable; sin embargo,

$$P^n = \begin{bmatrix} 46 & -1 & \sqrt{3} \\ 25 & 1 & 0 \\ 20 & 0 & -\sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{20} & -\frac{\sqrt{3}}{20} \\ 0 & \frac{\sqrt{3}}{20} & \frac{1}{20} \end{bmatrix}^n \begin{bmatrix} 46 & -1 & \sqrt{3} \\ 25 & 1 & 0 \\ 20 & 0 & -\sqrt{3} \end{bmatrix}^{-1}.$$

Luego,

$$\lim_{n\to\infty} P^n = \frac{1}{273}\begin{bmatrix} 46 & -1 & \sqrt{3} \\ 25 & 1 & 0 \\ 20 & 0 & -\sqrt{3} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{20} & \frac{\sqrt{3}}{20} \\ 0 & -\frac{\sqrt{3}}{20} & \frac{1}{20} \end{bmatrix}^n \begin{bmatrix} 3 & 3 & 3 \\ -75 & 198 & -75 \\ 20\sqrt{3} & 20\sqrt{3} & -71\sqrt{3} \end{bmatrix}.$$

Aquí, las potencias de la forma $3n$ cumplen que

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{20} & -\frac{\sqrt{3}}{20} \\ 0 & \frac{\sqrt{3}}{20} & \frac{1}{20} \end{bmatrix}^{3n} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{(-1)^n}{10^{3n}} & 0 \\ 0 & 0 & \frac{(-1)^n}{10^{3n}} \end{bmatrix},$$

y como $\left(\frac{-1}{10^3}\right)^n$ tiende a cero cuando $n \to \infty$,

$$\lim_{n\to\infty} P^{3n} = \frac{1}{91}\begin{bmatrix} 46 & 46 & 46 \\ 25 & 25 & 25 \\ 20 & 20 & 20 \end{bmatrix}.$$

Nótese que $\lim_{n\to\infty} P^n$ da como resultado una matriz con todas sus columnas iguales al vector de probabilidad correspondiente al vector propio asociado al valor propio $\lambda = 1$. Para convertir el vector propio $\vec{v}_1 = \begin{bmatrix} 46 \\ 25 \\ 20 \end{bmatrix}$, en un vector de probabilidad, se divide cada una de sus componentes por la suma de todos sus elementos, es decir

$$\vec{v} = \frac{1}{91}\vec{v}_1 = \frac{1}{91}\begin{bmatrix} 46 \\ 25 \\ 20 \end{bmatrix}.$$

Por lo tanto, el vector de estados a largo plazo es

$$\lim_{n\to\infty} X^{(n)} = \lim_{n\to\infty} P^n X^{(0)} = \frac{1}{91}\begin{bmatrix} 46 & 46 & 46 \\ 25 & 25 & 25 \\ 20 & 20 & 20 \end{bmatrix}\begin{bmatrix} \frac{1}{5} \\ \frac{1}{5} \\ \frac{3}{5} \end{bmatrix} = \begin{bmatrix} \frac{46}{91} \\ \frac{25}{91} \\ \frac{20}{91} \end{bmatrix}$$

Entonces, se puede decir que a largo plazo la marca A tendrá el control de cerca del $\frac{46}{91} \approx 50.55\,\%$ del mercado, la marca B el $\frac{25}{91} \approx 27.47\,\%$ del mercado y la otra marca el $\frac{20}{91} \approx 21.98\,\%$ del mercado.

Ejemplo 10.2 *Supóngase que el clima en cierta ciudad es bueno, regular o malo. Si el clima es bueno hoy, será bueno mañana con una probabilidad de* 0.60*, será regular con una probabilidad de* 0.20 *y será malo con una probabilidad de* 0.20*. Si el clima es regular hoy, será bueno, regular o malo con probabilidades de* $0.25, 0.50, 0.25$*, respectivamente. Ahora, si el clima es malo hoy, mañana será bueno, regular o malo con probabilidades de* $0.25, 0.25, 0.50$*. ¿A largo plazo cuál será el porcentaje de días buenos, regulares y malos?*

Solución.

Se aborda el problema por medio de las cadenas de Markov con

$$P = \begin{array}{c} \\ B \\ R \\ M \end{array} \begin{array}{c} \begin{array}{ccc} B & R & M \end{array} \\ \begin{bmatrix} 0.60 & 0.20 & 0.20 \\ 0.25 & 0.50 & 0.25 \\ 0.25 & 0.25 & 0.50 \end{bmatrix} \end{array}.$$

Como P es estocástica por filas, entonces se transpone

$$P^t = \begin{bmatrix} \frac{3}{5} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{5} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

En este caso, los valores propios de P^t son $\lambda_1 = 1$, $\lambda_2 = \frac{7}{20}$ y $\lambda_3 = \frac{1}{4}$, y los vectores propios correspondientes son $\vec{v}_1 = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, respectivamente.

Luego, la diagonalización de P^t da como resultado

$$\left(P^t\right)^n = \begin{bmatrix} 5 & -2 & 0 \\ 4 & 1 & -1 \\ 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{7}{20} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}^n \begin{bmatrix} 5 & -2 & 0 \\ 4 & 1 & -1 \\ 4 & 1 & 1 \end{bmatrix}^{-1}.$$

Dado que $\left(\frac{7}{20}\right)^n$ y $\left(\frac{1}{4}\right)^n$ tienden a cero cuando $n \to \infty$,

$$\lim_{n\to\infty} \left(P^t\right)^n = \frac{1}{26} \begin{bmatrix} 5 & -2 & 0 \\ 4 & 1 & -1 \\ 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 & 2 \\ -8 & 5 & 5 \\ 0 & -13 & 13 \end{bmatrix} = \frac{1}{13} \begin{bmatrix} 5 & 5 & 5 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

Nótese que $\lim_{n\to\infty}\left(P^t\right)^n$ da como resultado una matriz con todas sus columnas iguales al vector de probabilidad correspondiente al vector propio asociado al valor propio $\lambda = 1$. Para convertir el vector propio $\vec{v}_1 = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, en un vector de probabilidad, se divide cada una de sus componentes por la suma de todos sus elementos, es decir

$$\vec{v} = \frac{1}{13}\vec{v}_1 = \frac{1}{13}\begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}.$$

Entonces, se puede decir que a la larga el clima será bueno el $\frac{5}{13} \approx 38.46\,\%$ de los días, regular el $\frac{4}{13} \approx 30.77\,\%$ de los días y malo el $\frac{4}{13} \approx 30.77\,\%$ de los días.

10.2 Modelos genéticos

La relación entre las matemáticas y la biología forma parte de un problema antiguo en la historia de las ciencias. Esta sección se ha dedicado a una de las aplicaciones de métodos de modelación matemática en genética. Para desarrollar este tipo de aplicación únicamente es requisito conocer el proceso para diagonalizar una matriz desde el punto de vista matemático, ya que también se tendrán que manejar términos propios utilizados en el desarrollo de la genética. Para una mejor comprensión, se darán a conocer algunas nociones básicas de este tema. Se comienza con una reseña histórica muy breve de cómo inició la genética y de cómo desde el principio estuvo muy ligada a la estadística; además, se da una explicación sencilla de qué es la genética en cuanto a lo que al lector le interesa.

La ciencia de la genética nació en 1900, cuando varios investigadores de la reproducción de las plantas descubrieron el trabajo del monje austríaco Gregor Mendel, cuya publicación había sido ignorada en la práctica. En Mendel (1865) se presenta el resultado de cerca de 10 años de observaciones minuciosas con la planta del guisante (o chícharo) y se describen los patrones de la herencia en función de siete pares de rasgos contrastantes que aparecían en siete variedades diferentes de esta planta. En su trabajo observó que los caracteres se heredaban como unidades separadas y cada

una de ellas lo hacía de forma independiente con respecto a las otras. Esto le permitió señalar que cada progenitor tiene pares de unidades, pero que solo aporta una unidad de cada pareja a su descendiente. Su investigación logró expresarla matemáticamente mediante las leyes de la probabilidad y así predijo los resultados de los cruces genéticos, datos que se escriben en fracciones o porcentajes. Más tarde, las unidades descritas por Mendel recibieron el nombre de genes.

La observación obtenida al cruzar dos plantas puras con diferentes caracteres llevó a Mendel a deducir que existía un rasgo más fuerte, al que llamó *dominante*, y un rasgo más débil o que aparentemente desaparece al que le dio el nombre de *recesivo*. Estos dos conceptos de rasgo dominante y recesivo, aunque muy fáciles de comprender, son de vital importancia a la hora de desarrollar la parte matemática de esta rama.

En la historia de la biología este ha sido uno de los experimentos más extensos que ha realizado un solo autor. La recepción que tuvo esta publicación fue prácticamente nula entre la comunidad científica de su época. Después de casi cuatro décadas, las leyes de Mendel fueron redescubiertas. A partir de entonces, comenzó el desarrollo impetuoso de la genética. Aun cuando sea un simplismo, que la comunidad científica ignorara las leyes de Mendel, ha costado cuarenta años de retraso a la biotecnología moderna.

Un gen particular puede ocurrir en varias formas o *alelos*. Para simplificar, consideraremos un gen con dos alelos, los genetistas denotan los caracteres dominantes con letras mayúsculas y los caracteres recesivos, con minúsculas. De esta manera, los alelos serán A y a.

10.2.1 Herencia autosómica

En esta sección se considera la herencia como *autosómica*, esto quiere decir que un individuo hereda un gen de cada uno de los genes de sus padres, formando así su propio par. Hasta donde se sabe, es el azar el que determina cuál de los dos genes de un progenitor pasa a su descendiente. Si fuera posible clasificar los individuos de una población de una especie dada en cuanto a los genotipos AA, Aa y aa (téngase en cuenta que el genotipo Aa es igual que el aA), se podrían determinar las proporciones de los alelos en la población. Esto no sería factible si, por ejemplo, no se pudieran distinguir AA de Aa.

Para $n = 0, 1, 2, \ldots$, se establecen las siguientes proporciones:

p_n : Proporción del genotipo AA que hay en la generación de orden n.

q_n : Proporción del genotipo Aa que hay en la generación de orden n.

r_n : Proporción del genotipo aa que hay en la generación de orden n.

Si se supone que se pueden determinar esas proporciones, nótese que se debe tener

$$p_n + q_n + r_n = 1. \tag{10.4}$$

Entonces, las proporciones u y v de los dos alelos A y a en la población satisfacen las ecuaciones

$$u = p_n + \frac{1}{2}q_n \qquad \text{y} \qquad v = \frac{1}{2}q_n + r_n. \tag{10.5}$$

Aquí, se usó el hecho de que los alelos A y a constituyen el 100 % del genotipo AA (con proporción p_n) y el 50 % del genotipo Aa, y similarmente para los alelos. Si se supone que los genotipos ocurren en las mismas proporciones entre los machos y las hembras, entonces u y v representan (en toda la población) las probabilidades de que el gen sea A o a, respectivamente.

Ejemplo 10.3 *En una población, la distribución de genotipos en la n-ésima generación es de* 50 % *de AA,* 30 % *de Aa y* 20 % *de aa. ¿Qué proporciones de los genes en esta población son A y a?*

Solución.

En este ejemplo, $p_n = 0.50$, $q_n = 0.30$ y $r_n = 0.20$. Por lo tanto,

$$u = 0.50 + \frac{1}{2}(0.30) = 0.65 \qquad \text{y} \qquad v = 0.15 + 20 = 0.35.$$

Es decir, que de la "población" de genes el 65 % es de alelo A y el 35 % es de alelo a.

Con frecuencia, es interesante el problema inverso al de la determinación de las proporciones de los genotipos cuando se conocen las proporciones de los alelos. En general, este problema no tiene solución única. El sistema de ecuaciones dado en (10.5) se reduce a una ecuación de dos incógnitas, $u = p_n + (1/2)q_n$. Para obtener una segunda ecuación independiente, supondremos apareamiento aleatorio. Esto quiere decir que la probabilidad de que un individuo dado se aparee con otro individuo no depende del genotipo de este último. En muchos casos, esta es una suposición correcta. En otros no; por ejemplo, se sabe que la gente alta tiende a casarse con gente alta y, por lo tanto, la característica de la estatura

en los humanos no se puede analizar de esta manera. Por otro lado, se ha demostrado que la suposición de apareo aleatorio se aplica a la característica de los tipos de sangre humana. La mayoría de los individuos escogen su cónyuge sin preocuparse por su tipo de sangre.

De nuevo, supóngase que u y v son las proporciones de los alelos A y a entre los machos y entre las hembras. Entonces, si suponemos que la población es grande, la probabilidad de que la descendencia reciba el alelo A de los dos padres es u^2. De manera similar, las probabilidades de los genotipos AA y aa son $2uv$ y v^2, respectivamente. El término $2uv$ viene del hecho de que los alelos Aa y aA son el mismo, hecho que ya se había enunciado. Este resultado conduce al siguiente teorema, descubierto de manera independiente por Hardy y Weinberg en 1908.

Teorema 10.6 Ley de Hardy–Weinberg

Supóngase que en una gran población de padres, los alelos A y a de un gen en particular se presentan en las proporciones u y $v = 1 - u$. Suponiendo que estas proporciones son las mismas para los machos y para las hembras y, además, que el apareo es aleatorio, la primera y todas las generaciones sucesivas se compondrán de los tres genotipos, AA, Aa y aa en las proporciones u^2, $2uv$ y v^2, respectivamente.

Demostración.

Como se ha visto, un individuo de la primera generación es de genotipo AA si sus dos padres contribuyen con los alelos A. Como la probabilidad es u de que cualquiera de los padres contribuya con un alelo A, la probabilidad del genotipo AA en la descendencia inmediata es de v^2. De manera análoga, las probabilidades de los genotipos Aa y aa son de $2uv$ y v^2, respectivamente. Esto implica que las proporciones p_1 y q_1 de los alelos A y a en la primera generación están dadas por

$$p_1 = u^2 + \frac{1}{2}(2uv) = u(u+v) = u$$

y

$$q_1 = \frac{1}{2}(2uv) + v^2 = v(u+v) = v.$$

Por lo tanto, las proporciones de los dos alelos no se afectan por la generación inicial. Esto continúa de generación en generación. Concluimos que, después de la generación inicial, las proporciones de los tres genotipos AA, Aa y aa permanecen constantes en u^2, $2uv$ y v^2. ■

Ejemplo 10.4 *El color de la semilla de chícharo está controlado por un par de genes. Los tres genotipos AA, Aa y aa se caracterizan por sus semillas de color rojo, rosa y blanco, respectivamente. Si se cultiva un campo al azar con* 60 % *de flores rojas y* 40 % *de flores blancas. ¿Qué proporciones de los tres genotipos estarán presentes en la cuarta generación?*

Solución

En este ejemplo, $u = 0.6$ y $v = 0.4$. Por la ley de Hardy–Weinberg, las proporciones de flores rojas, rosadas y blancas en la primera generación y en todas las subsecuentes son de u^2, $2uv$ y v^2, es decir, 0.36, 0.48 y 0.16, respectivamente. Nótese que la suposición de cultivo aleatorio equivale a la suposición de polinización aleatoria.

La ley de Hardy–Weinberg solo es válida cuando el apareamiento es aleatorio y cuando los tres genotipos son igualmente probables. En ciertos casos, es bastante difícil verificar que el apareo es aleatorio. Sin embargo, si las proporciones de los genotipos permanecen constantes durante varias generaciones y si satisfacen la ley de Hardy-Weinberg, esto se puede tomar como una fuerte evidencia de que el apareamiento es aleatorio. Así, como el conocimiento de que el apareo es aleatorio para los tipos de sangre humana, también algunas características para las plantas y los animales, se derivaron de las observaciones de las proporciones de los genotipos que cumplen esta ley.

10.2.2 Los cuadros de Punnett

Un cuadro de Punnett es una gráfica que muestra todas las combinaciones posibles de genes resultantes del cruce de dos organismos (de quienes los genes son conocidos). Se nombran cuadros de Punnett por el genetista inglés Reginald Punnett. Él descubrió algunos principios básicos de la

genética, incluso la unión del sexo y determinación del sexo. Además, trabajó con las características del color de las plumas de los pollos de manera separada para pollos machos y hembras.

Para ilustrar cómo se construye un cuadro de Punnett, se debe tener en cuenta que si uno de los padres es del genotipo Aa, entonces es igualmente probable que el descendiente herede de este progenitor el alelo A o el alelo a. Por otra parte, si uno de los padres es de genotipo aa y el otro es de Aa, el descendiente recibirá siempre un alelo a del progenitor de genotipo aa y un alelo A o a, con la misma probabilidad del progenitor de genotipo Aa. Así, el descendiente tiene la misma probabilidad de ser de genotipo AA o Aa. En la Tabla 10.1, se ubican las probabilidades de los posibles genotipos de los descendientes para todas las combinaciones posibles de los genotipos de los padres:

Genotipos	**Progenitores**					
Hijos	AA-AA	AA-Aa	AA-aa	Aa-Aa	Aa-aa	aa-aa
AA	1	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0
Aa	0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	0
aa	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	1

Tabla 10.1: Probabilidades de los posibles genotipos

Situaciones en el que el apareamiento no es aleatorio, se presentan frecuentemente en experimentos biológicos controlados. Un ejemplo evidente se da en la cría de caballos de carreras, en la que un ganador probado tiene gran demanda como semental. El ejemplo siguiente muestra una de las situaciones de apareamiento controlado.

Ejemplo 10.5 *Un agricultor tiene una gran población de plantas con cierta distribución de los tres posibles genotipos, AA, Aa y aa. Este hombre desea iniciar un programa de cultivos en el que todas las plantas de la población sean fecundadas por una planta del genotipo AA. Se quiere obtener la fórmula de la distribución de los tres posibles genotipos de la población, después de un cierto número de generaciones.*

Solución

Sean p_n, q_n y r_n las proporciones de los tres genotipos en la generación n. Luego, para $n = 0, 1, 2, \ldots$, se tiene que

$$p_n = p_{n-1} + \frac{1}{2}q_{n-1}, \qquad q_n = \frac{1}{2}q_{n-1} + r_{n-1}, \qquad r_n = 0. \tag{10.6}$$

Estas ecuaciones determinan la distribución de los genotipos en cada generación a partir de la distribución en la generación anterior y se lograron establecer por medio de la Tabla 10.1. El sistema (10.6) se puede expresar en notación matricial como

$$X^{(n)} = PX^{(n-1)} \qquad n = 1, 2, \ldots, \tag{10.7}$$

donde

$$X^{(n)} = \begin{bmatrix} p_n \\ q_n \\ r_n \end{bmatrix}, \quad X^{(n-1)} = \begin{bmatrix} p_{n-1} \\ q_{n-1} \\ r_{n-1} \end{bmatrix} \quad \text{y} \quad P = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Nótese que las columnas de la matriz P son iguales a las tres primeras columnas dadas en la Tabla 10.1.

La ecuación (10.7) implica

$$\begin{aligned} X^{(1)} &= PX^{(0)} \\ X^{(2)} &= PX^{(1)} = P\big(PX^{(0)}\big) = P^2X^{(0)} \\ X^{(3)} &= PX^{(2)} = P\big(P^2X^{(0)}\big) = P^3X^{(0)}, \end{aligned}$$

y en general,

$$X^{(n)} = P^nX^{(0)}. \tag{10.8}$$

Así, las proporciones de los genotipos futuros están completamente determinados por la matriz P y el vector $X^{(0)}$ de las proporciones iniciales.

Ahora, es fácil comprobar que los valores propios de P son $\lambda_1 = 1, \lambda_2 = \frac{1}{2}$ y $\lambda_3 = 0$, con vectores propios correspondientes:

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} \frac{1}{2} \\ -1 \\ \frac{1}{2} \end{bmatrix}.$$

Luego, P será diagonalizable por la matriz

$$C = \begin{bmatrix} 1 & -1 & \frac{1}{2} \\ 0 & 1 & -1 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}, \quad \text{ya que} \quad C^{-1}PC = D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Por último, como $P = CDC^{-1}$, se tiene que

$$P^n = (CDC^{-1})^n.$$

Este hecho no es desconocido para nosotros, pues

$$P^n = CD^nC^{-1}.$$

Determinar C^{-1} no tiene mayor inconveniente, después de un cálculo breve, se llega a

$$C^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix},$$

y como ya se halló C y C^{-1}, y como también sabemos que la matriz D es la matriz diagonal que contiene los valores propios asociados a P, se tiene que

$$D^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1}{2}\right)^n & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Como $\lim\limits_{n\to\infty} \left(\frac{1}{2}\right)^n \to 0$, se ve que D^n tiende a la matriz

$$\lim_{n\to\infty} D^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

de donde se tiene que

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 1 & -1 & \frac{1}{2} \\ 0 & 1 & -1 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Nótese que como P es estocástica por columnas, entonces $\lim\limits_{n\to\infty} P^n$ dio como resultado una matriz con todas sus columnas iguales al vector de probabilidad correspondiente al vector propio asociado al valor propio $\lambda = 1$.

Por otra parte,

$$\lim_{n\to\infty} X^{(n)} = \lim_{n\to\infty} P^n X^{(0)} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_0 \\ q_0 \\ r_0 \end{bmatrix} = \begin{bmatrix} p_0 + q_0 + r_0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

ya que $p_0 + q_0 + r_0 = 1$. Así, queda demostrado que a largo plazo todas las plantas serán genotipo AA.

10.3 Modelo de regresión lineal

El problema central de la inferencia estadística en una distribución bivariada es determinar la relación entre las variables y conocer de qué manera los cambios en una variable afectan la otra.

La variable que es la base de estimación es convencionalmente llamada *variable independiente* que se designa por X y la variable cuyo valor es estimado se llama *variable dependiente* la cual se designa por Y. La selección de las variables dependiente e independiente se hacen de acuerdo con lo conocido y con lo que se desee estimar. En este caso de dependencias entre variables, Y es una variable aleatoria pero X no lo es.

La naturaleza de la relación entre variables se establece a través del *análisis de regresión*. Esta es una técnica con la cual se establece la relación funcional entre las variables, de modo que permite predecir el valor que toma una variable en función del valor determinado de la otra. La regresión es generalmente clasificada en dos tipos: regresión simple y regresión múltiple o general.

La regresión simple hace referencia al estudio de las relaciones entre dos variables de las cuales una es independiente (X) y la otra es dependiente (Y).

La regresión múltiple comprende tres o más variables, una de las cuales es la variable dependiente que debe ser estimada con base en los valores de las otras variables que son las independientes.

Definición 10.8 Modelo

Es una relación entre dos o más variables cuantitativas, de tal forma que se pueda predecir una variable en función de otra u otras. En este punto es necesario distinguir entre dos tipos de relaciones:

1. *Una relación determinística o funcional es de la forma*

$$Y = f(X),$$

   donde X es la variable independiente y Y es la variable dependiente.

2. *Una relación estocástica o estadística no es una relación perfecta, es decir, no proporciona valores únicos de Y para valores determinados*

de X, pero se puede describir con precisión en términos probabilísticos.

En el análisis de regresión, se consideran relaciones del segundo tipo, no del primero.

Ejemplo 10.6 *La relación entre la variable aleatoria Y y la variable no aleatoria X puede ser expresada por:*

$$Y = \beta_0 \exp\big(\beta_1 X\big) + \varepsilon. \tag{10.9}$$

La ecuación (10.9) significa que para un valor dado de la variable X, el correspondiente de Y es la suma del valor $\beta_0 \exp\big(\beta_1 X\big)$ más una cantidad ε. Los parámetros son β_0 y β_1 y ε es la diferencia entre Y y el valor esperado de Y condicionada a un valor de X, es decir,

$$\varepsilon = Y - \mathbb{E}\big(Y|X\big).$$

Definición 10.9 Modelo lineal

Es una ecuación matemática que involucra variables aleatorias ligadas por parámetros y que es "lineal en los parámetros", y en algunas ocasiones en las variables aleatorias. La frase lineal en los parámetros significa que ningún parámetro en el modelo aparece como un exponente o es multiplicado (o dividido) por cualquier otro parámetro.

Ejemplo 10.7 *Cuáles de los siguientes modelos son lineales:*

$$(i)\ Y = \beta_0 + \beta_1 \frac{1}{X}. \qquad (ii)\ Y = \beta_0 + \beta_1 \sqrt{X}.$$

$$(iii)\ Y = \beta_0^2 + \beta_1 X.$$

Solución.

Los modelos dados en (i) y (ii) son lineales en los parámetros y el modelo dado en (iii) no es lineal en los parámetros, ya que β_0 no tiene exponente uno.

Definición 10.10 Modelo de regresión lineal simple

El modelo de la forma:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{10.10}$$

donde

y_i : *Es el valor de la variable respuesta en el i-ésimo ensayo.*

β_0, β_1 : *Son los coeficientes (o parámetros) de regresión que corresponden al intercepto y a la pendiente, respectivamente.*

x_i : *Es un valor conocido, el valor de la variable independiente en el i-ésimo ensayo.*

ε_i : *Es la componente aleatoria, y se conoce como error o perturbación.*

Se dice que es un modelo de regresión lineal simple. El nombre se debe al hecho de que es lineal tanto en los parámetros como en la variable independiente y simple porque solo se tiene una variable independiente.

La ecuación (10.10) es una expresión abreviada para el siguiente conjunto de n ecuaciones simultáneas:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ \vdots &= \quad \vdots \\ y_n &= \beta_0 + \beta_1 x_n + \varepsilon_n. \end{aligned} \tag{10.11}$$

El sistema de ecuaciones (10.11) puede escribirse en forma matricial como sigue

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{10.12}$$

$$\vec{Y} = \mathbf{X} \vec{\beta} + \vec{\varepsilon}.$$

Esta última representación permite formular la siguiente definición.

Definición 10.11 Modelo lineal general

Un modelo de la forma:

$$y_i = \sum_{j=0}^{p-1} \beta_j x_{ij} + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{10.13}$$

donde $x_{i0} = 1$. *El sistema de ecuaciones* (10.13) *puede escribirse en forma matricial como sigue:*

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}, \tag{10.14}$$

donde:

$\vec{Y}$: *Es un vector columna de tamaño* $n \times 1$, *de observaciones sobre la variable dependiente* Y.

$\mathbf{X}$: *Es una matriz de tamaño* $n \times p$, $p \leq n$, *de cantidades conocidas fijas, donde la primera columna es de unos y tiene rango igual a* $k \leq p$.

$\vec{\beta}$: *Es un vector columna de tamaño* $p \times 1$, *de parámetros desconocidos* $\beta_0, \beta_1, \ldots, \beta_{p-1}$.

$\vec{\varepsilon}$: *Es un vector aleatorio o perturbado de tamaño* $n \times 1$.

Se dice que es un modelo lineal general. Este modelo es de rango completo si el rango de $\mathbf{X}$ *es igual a* p.

10.3.1 Métodos de estimación de los parámetros del modelo

Para el modelo dado en (10.14), existen varios métodos de estimación de parámetros, entre los que se tiene:

- Mínimos cuadrados ordinarios ($M.C.O.$)
- Mínimos cuadrados generalizados o ponderados ($M.C.P.$)
- Máxima verosimilitud ($M.V.$)

En este capítulo se desarrollará una parte del método de $M.C.O.$; el lector que esté interesado en complementar dicho método y en los otros métodos, puede revisar el texto de Searle (1971).

10.3.1.1 Método de mínimos cuadrados ordinarios

El método de mínimos cuadrados ordinarios se atribuye a Carl Friedrich Gauss. Bajo ciertos supuestos, este método tiene algunas propiedades estadísticas muy atractivas que lo han convertido en uno de los más eficaces y populares del análisis de regresión. Son supuestos para su aplicación:

1. $\mathbb{E}\big[\vec{\varepsilon}\big] = \vec{0}$.
2. $\mathbb{E}\big[\vec{\varepsilon}\vec{\varepsilon}^{\,t}\big] = \sigma^2 I_n$.
3. La matriz $\mathbf{X}$ es no estocástica, es decir, consta de números fijos.
4. El rango de $\mathbf{X}$ es $\rho(X) = p$.
5. $\vec{\varepsilon}$ tiene una distribución normal multivariada, es decir,

$$\vec{\varepsilon} \sim N\big(\vec{0}, \sigma^2 I_n\big).$$

El supuesto 1 significa que el valor esperado del vector de perturbaciones (desviaciones) $\vec{\varepsilon}$, es decir, de cada uno de sus elementos, es cero. Más explícitamente, $\mathbb{E}\big[\vec{\varepsilon}\big] = \vec{0}$ significa que

$$\mathbb{E}\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix} = \begin{bmatrix}\mathbb{E}\big(\varepsilon_1\big)\\ \mathbb{E}\big(\varepsilon_2\big)\\ \vdots\\ \mathbb{E}\big(\varepsilon_n\big)\end{bmatrix} = \begin{bmatrix}0\\ 0\\ \vdots\\ 0\end{bmatrix}. \tag{10.15}$$

El supuesto 2 establece que las perturbaciones ε_i y ε_j no están correlacionadas y, además, que la varianza de ε_i para cada X_i (esto es, la varianza condicional de ε_i) es algún número positivo constante igual a σ^2, es

decir, representa el supuesto de *homocedasticidad*, o *igual* (homo) *dispersión* (cedasticidad) o *igual varianza*. Más explícitamente, $\mathbb{E}\big[\vec{\varepsilon}\vec{\varepsilon}^{\,t}\big] = \sigma^2 I_n$ significa que

$$\mathbb{E}\big[\vec{\varepsilon}\vec{\varepsilon}^{\,t}\big] = \mathbb{E}\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}\begin{bmatrix}\varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n\end{bmatrix} = \mathbb{E}\begin{bmatrix}\varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_n\\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \dots & \varepsilon_2\varepsilon_n\\ \vdots & \vdots & \ddots & \vdots\\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \dots & \varepsilon_n^2\end{bmatrix}.$$

Al aplicar el operador de valor esperado $\mathbb{E}$ a cada elemento de la matriz anterior, se obtiene

$$\mathbb{E}\big[\vec{\varepsilon}\vec{\varepsilon}^{\,t}\big] = \begin{bmatrix}\mathbb{E}\big(\varepsilon_1^2\big) & \mathbb{E}\big(\varepsilon_1\varepsilon_2\big) & \dots & \mathbb{E}\big(\varepsilon_1\varepsilon_n\big)\\ \mathbb{E}\big(\varepsilon_2\varepsilon_1\big) & \mathbb{E}\big(\varepsilon_2^2\big) & \dots & \mathbb{E}\big(\varepsilon_2\varepsilon_n\big)\\ \vdots & \vdots & \ddots & \vdots\\ \mathbb{E}\big(\varepsilon_n\varepsilon_1\big) & \mathbb{E}\big(\varepsilon_n\varepsilon_2\big) & \dots & \mathbb{E}\big(\varepsilon_n^2\big)\end{bmatrix}. \tag{10.16}$$

La matriz dada en (10.16) se denomina *matriz de varianza-covarianza* de las perturbaciones ε_i. Los elementos sobre la diagonal principal son las varianzas y los elementos por fuera de la diagonal principal son las covarianzas. Por definición,

$$\mathbb{E}\big\{\big[\varepsilon_i - \mathbb{E}\big(\varepsilon_i\big)\big]\big[\varepsilon_j - \mathbb{E}\big(\varepsilon_j\big)\big]\big\} = \begin{cases} Var\big(\varepsilon_i\big) & \text{si} \quad i = j\\ Cov\big(\varepsilon_i, \varepsilon_i\big) & \text{si} \quad i \neq j,\end{cases}$$

pero debido a los supuestos $\mathbb{E}\big(\varepsilon_i\big) = 0$, para cada i y $\mathbb{E}\big(\varepsilon_i\varepsilon_j\big) = 0$ si $i \neq j$, la matriz (10.16) se reduce a

$$\mathbb{E}\big[\vec{\varepsilon}\vec{\varepsilon}^{\,t}\big] = \begin{bmatrix}\sigma^2 & 0 & \dots & 0\\ 0 & \sigma^2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \sigma^2\end{bmatrix}. \tag{10.17}$$

El supuesto 3 estipula que la matriz $\mathbf{X}$ de tamaño $n \times p$ es no-estocástica, es decir, consiste en un conjunto de números fijos.

El supuesto 4 establece que la matriz $\mathbf{X}$ tiene rango columna completo, es decir, es igual a p, al número de columnas de la matriz. Esto significa que las columnas de la matriz $\mathbf{X}$ son linealmente independientes, es decir, no hay *relación lineal exacta* entre las variables X. En otras palabras, no hay *multicolinealidad*.

10.3.1.2 Forma operativa

Este método consiste en determinar los valores de los parámetros β_j de tal manera que la suma de cuadrados del error (SCE) sea mínima, es decir, se minimiza

$$SCE = \sum_{i=1}^{n} \varepsilon_i^2 = \vec{\varepsilon}^{\,t}\vec{\varepsilon} = \big(\vec{Y} - \mathbf{X}\vec{\beta}\big)^t\big(\vec{Y} - \mathbf{X}\vec{\beta}\big) \tag{10.18}$$

con respecto a $\vec{\beta}$. Las derivadas parciales de $\vec{\varepsilon}^{\,t}\vec{\varepsilon}$ con respecto a $\vec{\beta}$ dan

$$\frac{\partial(\varepsilon^t\varepsilon)}{\partial\vec{\beta}} = 2\mathbf{X}^t\mathbf{X}\hat{\beta} - 2\mathbf{X}^t\vec{Y}.$$

Si se iguala al vector nulo, se llega a las ecuaciones normales de la teoría de $M.C.O.$

$$\mathbf{X}^t\mathbf{X}\hat{\beta} = \mathbf{X}^t\vec{Y}. \tag{10.19}$$

Cuando $\mathbf{X}^t\mathbf{X}$ sea no singular, premultiplicando por $\big(\mathbf{X}^t\mathbf{X}\big)^{-1}$, se tiene

$$MCO(\beta) = \widehat{\beta} = \big(\mathbf{X}^t\mathbf{X}\big)^{-1}\mathbf{X}^t\vec{Y} = \mathbf{C}\vec{Y}, \tag{10.20}$$

donde $\mathbf{C} = \big(\mathbf{X}^t\mathbf{X}\big)^{-1}\mathbf{X}^t$. Otras estimaciones para el modelo (10.14) mediante el método de $M.C.O.$ son

$$\begin{aligned}
&\widehat{Y} = \mathbf{X}\hat{\beta} = \big[\mathbf{X}\big(\mathbf{X}^t\mathbf{X}\big)^{-1}\mathbf{X}^t\big]\vec{Y} = H\vec{Y}\\
&\widehat{\epsilon} = \vec{Y} - \hat{Y} = \vec{Y} - H\vec{Y} = \big(I_n - H\big)\vec{Y}\\
&SCE = \widehat{\epsilon}^t\widehat{\epsilon} = \big[\big(I_n - H\big)\vec{Y}\big]^t\big(I_n - H\big)\vec{Y} = \vec{Y}^t\big(I_n - H\big)\vec{Y} \qquad (10.21)\\
&SCT = \vec{Y}^t(I_n - \overline{J}_n)\vec{Y}\\
&SCR = SCT - SCE = \vec{Y}^t\big(H - \overline{J}_n\big)\vec{Y} \qquad (10.22)
\end{aligned}$$

La penúltima expresión se tiene del Ejemplo 5.5.

Obsérvese que la matriz $H = \mathbf{X}\big(\mathbf{X}^t\mathbf{X}\big)^{-1}\mathbf{X}^t$, la cual en la literatura estadística se conoce como *"Matriz Hat"*, determina muchos de los resultados de las estimaciones por $M.C.O.$ Por ejemplo, cuando premultiplica al vector de respuestas $\vec{Y}$, se obtienen las predicciones de la variable dependiente, por eso en algunos textos de estadística la denominan *Matriz de predicción* y a la matriz $I_n - H$ la llaman *Matriz residual*, puesto que al anteponérsele a la variable dependiente $\vec{Y}$, se obtienen los respectivos residuales.

10.3.1.3 Propiedades de los elementos de la matriz H

La matriz $H = [h_{ij}]$ de tamaño $n \times n$ cumple que

a) $h_{ii} = \sum_{j=1}^{n} h_{ij}^2 = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2$, ya que $H = H^t$ y $H = H^2$.

b) $0 < h_{ii} \leq 1$, si $i = 1, 2, \ldots, n$.

c) $-0.5 \leq h_{ij} \leq 0.5$, para $i \neq j$.

d) $\big(1 - h_{ii}\big)\big(1 - h_{jj}\big) - h_{ij}^2 \geq 0$.

e) $h_{ii}h_{jj} - h_{ij}^2 \geq 0$

f) Si $h_{ii} = 1$, entonces $h_{ij} = 0$, para todo $j \neq i$.

Si la matriz X de tamaño $n \times r$ es de rango r, entonces

g) $\sum_{i=1}^{n} h_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}^2 = r = \text{tr}(H)$.

h) $\sum_{i=1}^{n} h_{ij} = \sum_{j=1}^{n} h_{ij} = 1$.

Además, como $h_{ij} = x_i\big(\mathbf{X}^t\mathbf{X}\big)^{-1}x_j'$, h_{ii} está determinada por la localización de x_i en el espacio $\mathbf{X}$. Es decir, un pequeño (grande) valor de h_{ii} indica que x_i se encuentra cerca (lejos) de la masa de los otros puntos.

Ejemplo 10.8 *Ajuste el modelo de regresión:*

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i \tag{10.23}$$

al conjunto de datos hipotéticos de la siguiente tabla:

$\mathbf{x_1}$	$\mathbf{y}$	$\mathbf{x_1}$	$\mathbf{y}$	$\mathbf{x_1}$	$\mathbf{y}$
1	*-10*	*5*	*-2*	*9*	*6*
2	*-8*	*6*	*0*	*10*	*8*
3	*-6*	*7*	*2*	*11*	*10*
4	*-4*	*8*	*4*		

.

Solución.

Si se expresa el modelo dado en (10.23) en la forma matricial dada en (10.14), se tiene

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}, \tag{10.24}$$

donde $\mathbf{X} = \begin{bmatrix} X_0 & X_1 \end{bmatrix}$ y $X_0 = \mathbf{1}$. Al calcular $\left(\mathbf{X}^t\mathbf{X}\right)$, se tiene que

$$\left(\mathbf{X}^t\mathbf{X}\right) = \begin{bmatrix} 11 & 66 \\ 66 & 506 \end{bmatrix} = 11 \begin{bmatrix} 1 & 6 \\ 6 & 46 \end{bmatrix}.$$

Luego, las ecuaciones normales (10.19) son

$$11 \begin{bmatrix} 1 & 6 \\ 6 & 46 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 220 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} 1 & 6 \\ 6 & 46 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 20 \end{bmatrix}. \tag{10.25}$$

En el Ejemplo 7.8, se obtuvo que el $\kappa(A) \approx 218.9$, luego, existe multicolinealidad moderada, es decir que variaciones muy pequeñas en la varianza de la variable regresora X_1 produce cambios drásticos en las estimaciones de los parámetros.

Sin embargo, se puede establecer la solución de (10.25) como sigue:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 46 & -6 \\ -6 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 20 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} -120 \\ 20 \end{bmatrix} = \begin{bmatrix} -12 \\ 2 \end{bmatrix}. \tag{10.26}$$

Así, el modelo de regresión es

$$\widehat{y}_i = -12 + 2x_i, \qquad i = 1, 2, \ldots, 11.$$

10.4 Multicolinealidad

La *multicolinealidad* se refiere a la existencia de más de una relación lineal exacta. Inicialmente, este término significó la existencia de una relación "perfecta" o exacta entre algunas o todas las variables explicativas de un modelo de regresión. Para la regresión con p variables que incluye las variables explicativas $X_0, X_1, \ldots, X_p$ (donde $X_0 = 1$ para todas las observaciones que den cabida al término intercepto), se dice que existe una relación lineal exacta si se satisface la siguiente condición:

$$\alpha_0 X_0 + \alpha_1 X_1 + \ldots + \alpha_p X_p = \vec{0}, \qquad \alpha_i \in \mathbb{R}, \tag{10.27}$$

donde algunas constantes no son iguales a cero.

Sin embargo, el término multicolinealidad se utiliza también para el caso en el cual las variables $X_0, X_1, \ldots, X_p$ están intercorrelacionadas pero no en forma perfecta, de la siguiente manera:

$$\alpha_0 X_0 + \alpha_1 X_1 + \ldots + \alpha_p X_p + \nu_i = \vec{0}, \tag{10.28}$$

donde ν_i es un término de error estocástico.

En los textos de econometría se emplea como medida para detectar la multicolinealidad, el índice de condición, de la siguiente manera:

Si $0 \leq IC\big(\mathbf{X}^t\mathbf{X}\big) \leq 10$ no existe multicolinealidad,
$10 < IC\big(\mathbf{X}^t\mathbf{X}\big) \leq 30$ existe multicolinealidad entre moderada y fuerte,
$30 < IC\big(\mathbf{X}^t\mathbf{X}\big)$ existe multicolinealidad severa.

10.4.1 Soluciones al problema de la multicolinealidad

10.4.1.1 Regresión por componentes principales

Una solución que muchas veces se sugiere para el problema de la multicolinealidad es la regresión por componentes principales. Supongamos que se tiene un conjunto de p variables explicativas, $X_1, X_2, \ldots, X_p$. Entonces, se construyen funciones lineales de estas variables

$$Z_i = a_{i1}X_1^* + a_{i2}X_2^* + \ldots + a_{ip}X_p^*, \qquad i = 1, 2, \ldots, p, \tag{10.29}$$

con $X_i^* = \frac{X_i - \overline{X_i}}{S_{X_i}}$, de tal manera que un grupo m $\big(m < p\big)$ de las variables $Z_1, Z_2, \ldots, Z_p$ contengan aproximadamente la misma información que $X_1, X_2, \ldots, X_p$. Las variables $Z_1, Z_2, \ldots, Z_p$ se buscan de manera que

(i) $Cov\big(Z_i, Z_j\big) = 0$, $i, j = 1, 2, \ldots, p$, $i \neq j$.

(ii) $Var\big(Z_1\big) \geq Var\big(Z_2\big) \geq \ldots \geq Var\big(Z_p\big)$.

(iii) $Z_i = \vec{a}_i^t \mathbf{X}_*$ para $i = 1, 2, \ldots, p$ y está sujeto a la condición de que

$$\|\vec{a}_i\|_2 = \vec{a}_i^t \vec{a}_i = 1, \tag{10.30}$$

donde $\vec{a}_i^t = \big(a_{i1}, a_{i2}, \ldots, a_{ip}\big)$ y $\mathbf{X}_* = \big[X_1^* \quad X_2^* \quad \ldots \quad X_p^*\big]^t$.

La matriz de covarianzas de $\mathbf{X}_*$ es $S = \mathbb{E}\big(\mathbf{X}_*\mathbf{X}_*^t\big)$, la cual es una matriz simétrica definida (o semidefinida) positiva y de orden p. Como

$$Var\big(Z_i\big) = \mathbb{E}\big(Z_i {Z_i}^t\big) = \mathbb{E}\big(\vec{a}_i^t\mathbf{X}_*\mathbf{X}_*^t\vec{a}_i\big) = \vec{a}_i^t\mathbb{E}\big(\mathbf{X}_*\mathbf{X}_*^t\big)\vec{a}_i = \vec{a}_i^t S\vec{a}_i,$$

para hallar Z_1, se necesita conocer el vector de coeficientes $\vec{a}_1$, y para ello, puesto que la varianza de Z_1 debe ser mayor que la varianza de las restantes componentes, habrá que resolver el problema de optimización

$$\begin{aligned} &\max F\big(\vec{a}_1\big) = \vec{a}_1^t S\vec{a}_1, \\ &\vec{a}_1^t\vec{a}_1 = 1. \end{aligned}$$

Por lo tanto, consideremos la maximización de la forma cuadrática $\vec{a}_1^t S\vec{a}_1$, sujeta a la condición (10.30). Si se introduce el multiplicador de Lagrange δ, se maximiza

$$\vec{a}_1^t S\vec{a}_1 - \delta\big(\vec{a}_1^t\vec{a}_1 - 1\big).$$

Al diferenciar respecto a $\vec{a}_1$, δ e igualar a cero las derivadas, se obtiene

$$2S\vec{a}_1 - 2\delta\vec{a}_1 = \vec{0} \qquad \text{y} \qquad \vec{a}_1^t\vec{a}_1 - 1 = 0.$$

Por lo tanto,

$$S\vec{a}_1 = \delta\vec{a}_1 \qquad \text{y} \qquad \vec{a}_1^t\vec{a}_1 = 1.$$

Este sistema tiene como soluciones todos los vectores propios de la matriz S de norma 1 asociados a cada uno de los valores propios de S.

Sean $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ los valores propios de S (la positividad estricta de los valores propios λ_i está garantizada si S es definida positiva) y $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ los correspondientes vectores propios de S normalizados. Entonces, los puntos estacionarios del problema son

$$\vec{a}_1^{\,j} = \vec{v}_j, \qquad j = 1, 2, \ldots, p,$$

con multiplicadores de Lagrange asociados

$$\delta_j = \lambda_j, \qquad j = 1, 2, \ldots, p.$$

Entre todos estos puntos estacionarios, el máximo se alcanza en el que coincide con el vector propio de S correspondiente al valor propio dominante[1].

Además, como $\vec{v}_i$ es un vector propio normalizado de S, la forma cuadrática $\vec{a}_j^t S\vec{a}_j = \lambda_j$, de lo cual se deduce que

$$Var\big[Z_j\big] = \lambda_j.$$

[1] De hecho, hay que considerar $\lambda_1 > \lambda_2$, ya que si $\lambda_1 = \lambda_2$, entonces los vectores propios $\vec{v}_1$ y $\vec{v}_2$ asociados son ambos solución del problema del máximo.

10.4.1.2 Propiedades de los componentes

Los componentes principales tienen las siguientes propiedades:

(a) La suma de los primeros k valores propios dividida por la suma de todos los valores propios, es decir,

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p}$$

representa la "proporción de la variación total" explicada por las primeras k componentes principales.

(b) Las componentes principales de un vector aleatorio son invariantes a los escalares.

(c) Si la matriz de covarianza S tiene rango $r < p$, entonces la variación total de S puede ser explicada enteramente por las primeras r componentes principales.

(d) El subespacio vectorial formado por las primeras k componentes principales $1 \leq k \leq p$ tienen la desviación cuadrática media más pequeña de las variables de la población (o muestra) que cualquier otro subespacio k-dimensional.

(e) Como un caso especial de (d), para $k = p - 1$, el plano perpendicular a las últimas componentes principales tienen la desviación cuadrática media más pequeña de las variables de la población (o muestra) que cualquier otro plano.

Ejemplo 10.9 *Ajuste el modelo de regresión:*

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \tag{10.31}$$

al conjunto de datos hipotéticos de la siguiente tabla:

$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{y}$	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{y}$	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{y}$
1	*1*	*-10*	*5*	*9*	*-2*	*9*	*17*	*6*
2	*3*	*-8*	*6*	*11*	*0*	*10*	*19*	*8*
3	*5*	*-6*	*7*	*13*	*2*	*11*	*21*	*10*
4	*7*	*-4*	*8*	*15*	*4*			

.

Solución.

Si se expresa el modelo dado en (10.31) en forma matricial, se tiene

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}, \tag{10.32}$$

donde

$$\mathbf{X} = [X_0 \;\; X_1 \;\; X_2]\,.$$

Aquí, $X_0 = \mathbf{1}$ para todas las observaciones que den cabida al término intercepto, de modo que las ecuaciones normales son

$$\big(\mathbf{X}^t\mathbf{X}\big)\hat{\beta} = \mathbf{X}^t\vec{Y}. \tag{10.33}$$

Al calcular $\big(\mathbf{X}^t\mathbf{X}\big)$, se tiene que

$$\big(\mathbf{X}^t\mathbf{X}\big) = \begin{bmatrix} 11 & 66 & 121 \\ 66 & 506 & 946 \\ 121 & 946 & 1771 \end{bmatrix} = 11 \begin{bmatrix} 1 & 6 & 11 \\ 6 & 46 & 86 \\ 11 & 86 & 161 \end{bmatrix}.$$

Si se reescribe (10.33), se llega a

$$11 \begin{bmatrix} 1 & 6 & 11 \\ 6 & 46 & 86 \\ 11 & 86 & 161 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 220 \\ 440 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} 1 & 6 & 11 \\ 6 & 46 & 86 \\ 11 & 86 & 161 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 20 \\ 40 \end{bmatrix}. \tag{10.34}$$

Primero se determina si el sistema de ecuaciones obtenido en (10.34) es estable. Para ello, se debe calcular el número de condición dado en (7.6), por lo tanto, se necesitan los valores propios de la matriz $\big(\mathbf{X}^t\mathbf{X}\big)$,

$$p_{\left(\mathbf{X}^t\mathbf{X}\right)}(\lambda) = -\lambda^3 + 2288\lambda^2 - 7260\lambda = -\lambda\big(\lambda^2 - 2288\lambda + 7260\big),$$

de donde los valores propios son

$$\lambda_1 = 11\big(104 + 2\sqrt{2689}\big), \quad \lambda_2 = 11\big(104 - 2\sqrt{2689}\big) \quad \text{y} \quad \lambda_3 = 0.$$

Luego, el número de condición es muy grande, ya que

$$\kappa\big(\mathbf{X}^t\mathbf{X}\big) = \frac{\lambda_1}{\lambda_3}.$$

En este caso, se dice que existe multicolinealidad severa, es decir, variaciones muy pequeñas en las varianzas y las covarianzas de las variables regresoras X_i producen cambios drásticos en las estimaciones de los parámetros.

En el Ejemplo 9.18, se concluyó que el sistema dado en (10.34) era consistente indeterminado, y se obtuvo la siguiente solución:

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} = \frac{1}{3}\begin{bmatrix}\text{-}28\\ \text{-}10\\ 8\end{bmatrix}. \tag{10.35}$$

Como el sistema es consistente indeterminado, se utiliza el análisis de componentes principales para determinar el coeficiente β_i más significativo, y para establecer dicho coeficiente, se construye la matriz:

$$\Sigma = \frac{1}{n-1}\mathbf{X}^t\big(I_n - \overline{J}_n\big)\mathbf{X}.$$

Al efectuar los productos descritos anteriormente, se llega a

$$\Sigma = \begin{bmatrix}11 & 22\\ 22 & 44\end{bmatrix},$$

y con los elementos de $\Sigma = [\sigma_{ij}]$ se forma la matriz $S = [s_{ij}]$, donde

$$s_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

Luego,

$$S = \begin{bmatrix}\frac{11}{11} & \frac{22}{\sqrt{(11)(44)}}\\ \frac{22}{\sqrt{(11)(44)}} & \frac{44}{44}\end{bmatrix} = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}.$$

Los valores propios de la matriz S son $\lambda_1 = 2$ y $\lambda_2 = 0$, y sus respectivos vectores propios normalizados son $\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$ y $\vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}\text{-}1\\1\end{bmatrix}$.
Por lo tanto, las componentes principales son:

$$Z_1 = \frac{1}{\sqrt{2}}X_1^* + \frac{1}{\sqrt{2}}X_2^* = \frac{1}{\sqrt{2}}\big(X_1^* + X_2^*\big),$$
$$Z_2 = -\frac{1}{\sqrt{2}}X_1^* + \frac{1}{\sqrt{2}}X_2^* = \frac{1}{\sqrt{2}}\big(X_2^* - X_1^*\big),$$

y dado que

$$\mathbb{E}\big[Z_i\big] = 0 \qquad \text{y} \qquad Var\big[Z_i\big] = \lambda_i,$$

entonces la componente principal Z_2 tiene media 0 y varianza cero. El último resultado también se puede obtener de la definición de varianza, así:

$$\begin{aligned} Var\big[Z_2\big] &= Var\left[\frac{1}{\sqrt{2}}\big(X_2^* - X_1^*\big)\right] = \frac{1}{2}Var\big(X_2^* - X_1^*\big) \\ &= \frac{1}{2}\big[Var\big(X_2^*\big) + Var\big(X_1^*\big) - 2Cov\big(X_1^*, X_2^*\big)\big] \\ &= \frac{1}{2}\big[1 + 1 - 2\big] = 0 = \lambda_2. \end{aligned}$$

$Var\big[Z_2\big] = 0$ identifica la función lineal como el origen de la multicolinealidad. Luego, es posible decir que $Z_2 \simeq 0$, lo cual da que $X_1^* \simeq X_2^*$. Si se realiza la regresión de X_2 sobre X_1, es decir al expresar,

$$x_{2i} = \alpha_0 + \alpha_1 x_{1i},$$

se obtiene que

$$x_{2i} = -1 + 2x_{1i}.$$

Como existe una relación exacta entre X_2 y X_1, no es posible estimar por separado los coeficientes de X_1 y X_2. Por lo tanto, si en la ecuación original (10.31) se sustituye x_{2i} en términos de x_{1i}, se obtiene

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2\big(-1 + 2x_{1i}\big) + \varepsilon_i \\ &= \big(\beta_0 - \beta_2\big) + \big(\beta_1 + 2\beta_2\big)x_{1i} + \varepsilon_i. \end{aligned}$$

Esto da lo que en estadística se denomina *funciones lineales estimables* de $\hat{\beta}$, que son

$$\hat{\beta}_0 - \hat{\beta}_2 \qquad \text{y} \qquad \hat{\beta}_1 + 2\hat{\beta}_2. \tag{10.36}$$

Finalmente, en el Ejemplo 10.8 se obtuvo la regresión de y sobre x_1, lo cual dio como resultado que

$$y_i = -12 + 2x_{1i},$$

es decir,

$$\hat{\beta}_0 - \hat{\beta}_2 = -12 \qquad \text{y} \qquad \hat{\beta}_1 + 2\hat{\beta}_2 = 2.$$

Este sistema expresado matricialmente queda

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} -12 \\ 2 \end{bmatrix}. \tag{10.37}$$

Luego, una solución sería

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} = \begin{bmatrix}\text{-}12\\ 2\\ 0\end{bmatrix}. \tag{10.38}$$

Por otra parte, si en la ecuación (10.31) se sustituye x_{1i} en términos de x_{2i}, se llega a

$$\begin{aligned} y_i &= \beta_0 + \beta_1\left(\frac{x_{2i}+1}{2}\right) + \beta_2 x_{2i} + \varepsilon_i \\ &= \left(\beta_0 + \frac{1}{2}\beta_1\right) + \left(\frac{1}{2}\beta_1 + \beta_2\right)x_{2i} + \varepsilon_i. \end{aligned}$$

Luego las nuevas *funciones lineales estimables* de $\hat{\beta}$ son

$$\hat{\beta}_0 + \frac{1}{2}\hat{\beta}_1 \qquad \text{y} \qquad \frac{1}{2}\hat{\beta}_1 + \hat{\beta}_2. \tag{10.39}$$

En este caso, la regresión de y sobre x_2 da como resultado que

$$y_i = -11 + x_{2i},$$

es decir,

$$\hat{\beta}_0 + \frac{1}{2}\hat{\beta}_1 = -11 \qquad \text{y} \qquad \frac{1}{2}\hat{\beta}_1 + \hat{\beta}_2 = 1.$$

Este sistema expresado matricialmente queda

$$\begin{bmatrix} 1 & 0.5 & 0 \\ 0 & 0.5 & 1 \end{bmatrix} \begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} = \begin{bmatrix}\text{-}11\\ 1\end{bmatrix}. \tag{10.40}$$

Una solución sería

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} = \begin{bmatrix}\text{-}11\\ 0\\ 1\end{bmatrix}. \tag{10.41}$$

Dado que se obtuvieron dos soluciones distintas (10.38) y (10.41) para $\hat{\beta}$, entonces se usa la inversa generalizada para resolver el sistema (10.37), y se obtiene que la estimación de $\vec{\beta}$ viene dada por

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} = \frac{1}{6}\begin{bmatrix} 5 & 2 \\ 2 & 2 \\ \text{-}1 & 2 \end{bmatrix}\begin{bmatrix}\text{-}12\\ 2\end{bmatrix} = \frac{1}{6}\begin{bmatrix}\text{-}56\\ \text{-}20\\ 16\end{bmatrix}. \tag{10.42}$$

Nótese que este último vector coincide con el obtenido en (10.35).

Como lo indica este ejemplo, la multicolinealidad revela que no es posible estimar los coeficientes individuales de manera precisa, pero que sí se pueden calcular algunas funciones lineales de los parámetros.

10.5 Selección de carteras

En Markowitz (1952, 1959), se presentó una técnica de selección de carteras que llegó a ser llamada *la teoría moderna de cartera* (MPT, por su sigla en inglés). En esta teoría, se asume que los rendimientos esperados de los activos de la cartera tienen una distribución normal multivariante, supuesto análogo a las otras teorías que forman la base de las matemáticas financieras tradicionales. Este supuesto es una primera aproximación que da lugar a teorías analíticamente tratables. La optimización de las inversiones en el contexto media-varianza, se ha convertido en una de las herramientas con más aplicaciones en las decisiones financieras en sentido amplio. Para llevar a cabo la optimización de carteras, el inversor puede:

- Elegir un nivel de riesgo (volatilidad) dado, dependiendo de su aversión a este y buscar la cartera que mayor rendimiento le ofrezca.
- Seleccionar un rendimiento esperado y buscar la cartera que minimice el riesgo.

En esta sección se presenta el problema de decisión de selección de carteras. La cartera de mínima varianza es de interés particular, ya que las ponderaciones de los activos son determinadas a partir de la matriz de varianza-covarianza.

10.5.1 Formulación matemática

Suponga que hay m activos riesgosos, el rendimiento esperado del i-ésimo activo es denotado por $\mathbb{E}\big[r_i\big]$, la covarianza de los rendimientos entre el i-ésimo y j-ésimo activo es expresada como σ_{ij}, y la varianza del rendimiento del i-ésimo activo es denotada por $\sigma_{ii} = \sigma_i^2$. Además, se asume que ningún activo puede expresarse como una combinación lineal de los otros activos, entonces la matriz de varianza-covarianza de los rendimientos, $\Sigma = [\sigma_{ij}]$, es no singular. La frontera de todas las carteras factibles que pueden construirse con estos m activos se define como el lugar de carteras

factibles que tienen la variación más pequeña para un rendimiento esperado prescrito. Sean w_i $\{i = 1, \ldots, m\}$ constantes que determinan la ponderación de cada activo dentro de la cartera de tal forma que $\sum_{i=1}^{m} w_i = 1$, es decir, w_i se puede interpretar como la proporción de riqueza invertida en el i-ésimo activo. No se consideran ventas en corto y, por lo tanto, el rango de valores permitido para cada w_i se restringe al intervalo $[0, 1]$.

Dada la matriz de varianzas-covarianzas $\mathbf{\Sigma}$, la cartera de varianza mínima global ($GMVP$, por su sigla en inglés) es la cartera con ponderaciones $\vec{w} = [w_i]$ que soluciona el siguiente problema de optimización:

$$\min_{\vec{w}} \sigma_p^2 = \min_{\vec{w}} \vec{w}^t \mathbf{\Sigma} \vec{w}, \tag{10.43}$$

sujeto a la condición

$$\vec{w}^t \mathbf{1}_m = \mathbf{1}_m^t \vec{w} = 1, \tag{10.44}$$

donde $\vec{w}$ es el vector columna de tamaño $m \times 1$ cuyos elementos son las ponderaciones de la cartera, y al formar el lagrangeano, se tiene

$$\begin{aligned} \mathcal{L} &= \sigma_p^2 - 2\delta\big(\vec{w}^t \mathbf{1}_m - 1\big) \\ \frac{\partial}{\partial \vec{w}} \mathcal{L} &= 2\mathbf{\Sigma}\vec{w} - 2\delta \mathbf{1}_m, \end{aligned} \tag{10.45}$$

donde $\delta \in \mathbb{R}$ es el respectivo multiplicador de Lagrange. Al igualar a cero la ecuación (10.45), se tiene

$$\begin{aligned} \mathbf{\Sigma}\vec{w} &= \delta \mathbf{1}_m \\ \vec{w} &= \delta \mathbf{\Sigma}^{-1} \mathbf{1}_m, \end{aligned} \quad \text{o} \quad \begin{aligned} \vec{w}^t \mathbf{\Sigma} &= \delta \mathbf{1}_m^t \\ \vec{w}^t &= \delta \mathbf{1}_m^t \mathbf{\Sigma}^{-1}, \end{aligned}$$

y multiplicando esta expresión por la izquierda por $\mathbf{1}_m^t$, se obtiene que

$$\delta = \big(\mathbf{1}_m^t \mathbf{\Sigma}^{-1} \mathbf{1}_m\big)^{-1}. \tag{10.46}$$

Si el modelo fue correctamente descrito para la varianza-covarianza condicional, entonces se esperaría que la variación de cualquier cartera con ponderaciones $\vec{w}$ sea $\vec{w}^t \mathbf{\Sigma} \vec{w}$. Las variaciones de la cartera que son también relativamente pequeñas para la varianza pronosticada son indicio de exceso de correlación, mientras variaciones que son demasiado grandes indican subestimación de la correlación. Luego, al sustituir en (10.43), se llega a

$$\sigma_{Mp}^2 = \delta,$$

donde σ_{Mp}^2 denota la varianza mínima y bajo las condiciones dadas o restricciones consideradas, y está determinada de manera única y solo depende de la matriz $\mathbf{\Sigma}$.

10.5.2 Cartera con rentabilidad preestablecida

En este contexto, se supone que a los inversores solo les preocupa el rendimiento esperado *(su media)* y la varianza de dichos rendimientos *(riesgo)*. En este caso, la $GMVP$ estará conformada por las ponderaciones $\vec{w}$ que minimizan la varianza de la cartera para el nivel de rentabilidad esperado fijado por el inversor, por lo que el problema de optimización a resolver en el caso de m activos es el siguiente:

$$\min_{\vec{w}} \sigma_p^2 = \min_{\vec{w}} \vec{w}^t \Sigma \vec{w}, \tag{10.47}$$

pero sujeto a las restricciones

$$\begin{aligned} \vec{w}^t \mathbb{E}(\vec{r}) &= \mathbb{E}(\vec{r})^t \vec{w} = \widetilde{E} \\ \vec{w}^t \mathbf{1}_m &= \mathbf{1}_m^t \vec{w} = 1, \end{aligned}$$

donde $\mathbb{E}(r)$ es el rendimiento esperado, $\widetilde{E}$ es el rendimiento mínimo esperado de la inversión y $\vec{w}$ es el vector $m \times 1$ de ponderaciones. En forma matricial, estas restricciones quedan

$$\begin{bmatrix} \mathbb{E}(\vec{r})^t \\ \mathbf{1}_m^t \end{bmatrix} \vec{w} = \begin{bmatrix} \widetilde{E} \\ 1 \end{bmatrix}, \tag{10.48}$$

y al establecer el lagrangeano, se obtiene

$$\mathcal{L} = \vec{w}^t \mathbf{\Sigma} \vec{w} - 2\delta_1 \big(\vec{w}^t \mathbb{E}(\vec{r}) - \widetilde{E}\big) - 2\delta_2 \big(\vec{w}^t \mathbf{1}_m - 1\big),$$

donde δ_1, $\delta_2 \in \mathbb{R}$ son los multiplicadores de Lagrange. Las condiciones estándar de primer orden para un punto crítico son

$$\begin{aligned} \frac{\partial}{\partial \vec{w}} \mathcal{L} &= 2\mathbf{\Sigma}\vec{w} - 2\delta_1 \mathbb{E}(\vec{r}) - 2\delta_2 \mathbf{1}_m = \vec{0} \\ \frac{\partial}{\partial \delta_1} \mathcal{L} &= \vec{w}^t \mathbb{E}(\vec{r}) - \widetilde{E} = 0 \\ \frac{\partial}{\partial \delta_2} \mathcal{L} &= \vec{w}^t \mathbf{1}_m - 1 = 0, \end{aligned} \tag{10.49}$$

por lo tanto

$$\mathbf{\Sigma}\vec{w} = \delta_1 \mathbb{E}(\vec{r}) + \delta_2 \mathbf{1}_m \tag{10.50}$$

$$\vec{w} = \delta_1 \mathbf{\Sigma}^{-1} \mathbb{E}(\vec{r}) + \delta_2 \mathbf{\Sigma}^{-1} \mathbf{1}_m. \tag{10.51}$$

Para determinar los multiplicadores de Lagrange que satisfacen (10.49), se multiplica por la izquierda la ecuación (10.51) por el vector de rendimientos esperados transpuesto y se obtiene

$$\begin{aligned} \mathbb{E}(\vec{r})^t \vec{w} &= \mathbb{E}(\vec{r})^t \big(\delta_1 \mathbf{\Sigma}^{-1} \mathbb{E}(\vec{r}) + \delta_2 \mathbf{\Sigma}^{-1} \mathbf{1}_m\big) \\ \widetilde{E} &= \delta_1 B + \delta_2 A. \end{aligned} \tag{10.52}$$

Por otra parte, si se multiplica por la izquierda la ecuación (10.51) por un vector de unos transpuesto, se llega a

$$\begin{aligned} \mathbf{1}_m^t \vec{w} &= \mathbf{1}_m^t \big(\delta_1 \mathbf{\Sigma}^{-1} \mathbb{E}(\vec{r}) + \delta_2 \mathbf{\Sigma}^{-1} \mathbf{1}_m\big) \\ 1 &= \delta_1 A^t + \delta_2 C. \end{aligned} \tag{10.53}$$

Las expresiones (10.52) y (10.53) forman un sistema de ecuaciones, y al reescribirlo matricialmente, se tiene

$$\begin{bmatrix} B & A \\ A^t & C \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} \widetilde{E} \\ 1 \end{bmatrix},$$

y la solución sería

$$\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} B & A \\ A^t & C \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{E} \\ 1 \end{bmatrix} = \frac{1}{D} \begin{bmatrix} C & -A \\ -A^t & B \end{bmatrix} \begin{bmatrix} \widetilde{E} \\ 1 \end{bmatrix}, \tag{10.54}$$

donde las constantes A, B, C y D son definidas[2] como en Merton (1972)

$$\begin{array}{lll} A = \mathbb{E}(\vec{r})^t \mathbf{\Sigma}^{-1} \mathbf{1}_m, & & B = \mathbb{E}(\vec{r})^t \mathbf{\Sigma}^{-1} \mathbb{E}(\vec{r}), \\ C = \mathbf{1}_m^t \mathbf{\Sigma}^{-1} \mathbf{1}_m, & & D = B\,C - A^t\,A. \end{array} \tag{10.55}$$

Al multiplicar el lado izquierdo de la expresión (10.50) por $\vec{w}^t$ y sustituir (10.54), se obtiene que la varianza mínima de esta cartera es

$$\begin{aligned} \vec{w}^t \mathbf{\Sigma} \vec{w} &= \vec{w}^t \begin{bmatrix} \mathbb{E}(\vec{r}) & \mathbf{1}_m \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \\ &= \frac{1}{D} \begin{bmatrix} \widetilde{E} & 1 \end{bmatrix} \begin{bmatrix} C & -A \\ -A^t & B \end{bmatrix} \begin{bmatrix} \widetilde{E} \\ 1 \end{bmatrix}. \end{aligned} \tag{10.56}$$

[2] Las cantidades B y C son formas cuadráticas de $\mathbf{\Sigma}^{-1}$, luego son estrictamente positivas mientras que A es una forma bilineal simétrica de $\mathbf{\Sigma}^{-1}$.

Por otra parte[3], como $A = A^t$, sustituyendo las relaciones (10.55), se tiene que

$$\vec{w}^t \boldsymbol{\Sigma} \vec{w} \;=\; \frac{1}{D} \begin{bmatrix} \widetilde{E} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{1}_m^t \\ -\mathbb{E}(\vec{r})^t \end{bmatrix} \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{1}_m & -\mathbb{E}(\vec{r}) \end{bmatrix} \begin{bmatrix} \widetilde{E} \\ 1 \end{bmatrix}, \tag{10.57}$$

y al sustituir en (10.51) las expresiones obtenidas en (10.57), se llega a

$$\begin{aligned} \vec{w} \;&= \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbb{E}(\vec{r}) & \mathbf{1}_m \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \\ &= \frac{1}{D} \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbb{E}(\vec{r}) & \mathbf{1}_m \end{bmatrix} \begin{bmatrix} \mathbf{1}_m & -\mathbb{E}(\vec{r}) \end{bmatrix}^t \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{1}_m & -\mathbb{E}(\vec{r}) \end{bmatrix} \begin{bmatrix} \widetilde{E} \\ 1 \end{bmatrix} \\ &= \frac{1}{D} \boldsymbol{\Sigma}^{-1} K \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{1}_m & -\mathbb{E}(\vec{r}) \end{bmatrix} \begin{bmatrix} \widetilde{E} \\ 1 \end{bmatrix}, \end{aligned}$$

donde la matriz antisimétrica

$$K = \begin{bmatrix} \mathbb{E}(\vec{r}) & \mathbf{1}_m \end{bmatrix} \begin{bmatrix} \mathbf{1}_m & -\mathbb{E}(\vec{r}) \end{bmatrix}^t = \mathbb{E}(\vec{r}) \mathbf{1}_m^t - \mathbf{1}_m \mathbb{E}(\vec{r})^t$$

será igual a la matriz nula solo cuando todos los rendimientos esperados de los m activos sean iguales. Obsérvese que en estas ponderaciones ya no aparecen los multiplicadores de Lagrange; además, si se sustituye (10.48), se obtiene

$$\begin{aligned} \vec{w} \;&= D^{-1} \boldsymbol{\Sigma}^{-1} K \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{1}_m & -\mathbb{E}(\vec{r}) \end{bmatrix} \begin{bmatrix} \mathbb{E}(\vec{r}) & \mathbf{1}_m \end{bmatrix}^t \vec{w} \\ &= D^{-1} \boldsymbol{\Sigma}^{-1} K \boldsymbol{\Sigma}^{-1} K^t \vec{w} = -D^{-1} \big(\boldsymbol{\Sigma}^{-1} K \big)^2 \vec{w}. \end{aligned} \tag{10.58}$$

Reescribiendo esta última expresión, se llega a

$$\begin{gathered} \big[\big(\boldsymbol{\Sigma}^{-1} K \big)^2 - \big(-D \big) I_n \big] \vec{w} = \vec{0} \\ \big[\boldsymbol{\Sigma}^{-1} K - i \sqrt{D} I_n \big] \big[\boldsymbol{\Sigma}^{-1} K - \bar{i} \sqrt{D} I_n \big] \vec{w} = \vec{0}, \end{gathered}$$

donde i, $\bar{i}$ $\big(i.\bar{i} = 1\big)$ representan la unidad imaginaria y su conjugado complejo, respectivamente. Puesto que el vector de ponderaciones $\vec{w} \in \mathbb{R}^m$, este se construye realizando una combinación lineal de las correspondientes partes real e imaginaria del vector propio complejo, $\vec{u} \in \mathbb{C}^m$, asociado al valor propio $i\sqrt{D}$ de la matriz $\boldsymbol{\Sigma}^{-1} K$.

[3] Nótese que como estas dos formas cuadráticas son equivalentes y D es el determinante de la matriz simétrica asociada a la forma cuadrática del lado derecho, entonces es positivo.

Expresando el vector de ponderaciones como una combinación lineal de las correspondientes partes real e imaginaria del vector propio complejo $\vec{u} \in \mathbb{C}^m$, se tiene que

$$\vec{w} = \frac{1}{\Delta} W \begin{bmatrix} \mathbb{E}(\vec{r}) & -\mathbf{1}_m \end{bmatrix} \begin{bmatrix} 1 \\ \widetilde{E} \end{bmatrix}, \tag{10.59}$$

donde $\Delta = \mathbf{1}_m^t W \mathbb{E}(\vec{r})$ y

$$W = \begin{bmatrix} Re(\vec{u}) & Im(\vec{u}) \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} Re(\vec{u}) & Im(\vec{u}) \end{bmatrix}^t.$$

10.5.3 Cartera mínima con rentabilidad preestablecida

Al reescribir la forma cuadrática dada en (10.56), se tiene que

$$\begin{aligned} \vec{w}^t \boldsymbol{\Sigma} \vec{w} &= \frac{1}{D}\left(C\widetilde{E}^2 - 2A\widetilde{E} + B\right) \\ &= \frac{1}{D}\left[C\left(\widetilde{E} - \frac{A}{C}\right)^2 + \frac{D}{C}\right]. \end{aligned} \tag{10.60}$$

Luego, la frontera en el espacio media-varianza está determinada por la siguiente ecuación cuadrática:

$$C\, \vec{w}^t \boldsymbol{\Sigma} \vec{w} - \frac{C^2}{D}\left(\widetilde{E} - \frac{A}{C}\right)^2 = 1. \tag{10.61}$$

Esta cónica representa una hipérbola con centro en $\left(0, \frac{A}{C}\right)$, con vértices y focos sobre la recta horizontal $\widetilde{E}^{\text{mín}} = \frac{A}{C}$ y asíntotas dadas por

$$\widetilde{E} = \frac{A}{C} \pm \sqrt{\frac{D}{C}}\sigma, \qquad \text{con} \qquad \sigma = \sqrt{\vec{w}^t \boldsymbol{\Sigma} \vec{w}}.$$

Si se sustituye en (10.54) el valor mínimo $\widetilde{E}^{\text{mín}}$, se obtiene que

$$\delta_1 = 0 \qquad \text{y} \qquad \delta_2 = \frac{1}{C},$$

y al reemplazar en la ecuación (10.51), se reduce a

$$\vec{w} = \frac{1}{C} \boldsymbol{\Sigma}^{-1} \mathbf{1}_m. \tag{10.62}$$

Luego, al sustituir en (10.47), se llega a

$$\sigma_{Mp}^2 = \delta_2.$$

Nótese que estos resultados son los mismos que se obtuvieron en la sección anterior.

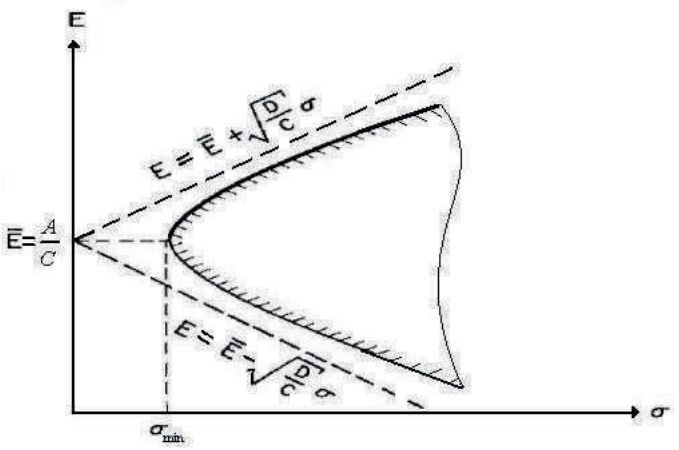

Figura 10.1: Frontera eficiente.

Ejemplo 10.10 Modelo de inversión

Considere tres categorías de inversión: acciones (x_1)*, bonos* (x_2) *y activos inmobiliarios* (x_3)*. Suponga que el rendimiento mínimo esperado es de* $\widetilde{E} = 9$ *y la matriz de varianza-covarianza es*

$$\mathbf{\Sigma} = \begin{bmatrix} \frac{9}{4} & -\frac{3}{25} & \frac{9}{20} \\ -\frac{3}{25} & \frac{16}{25} & \frac{42}{125} \\ \frac{9}{20} & \frac{42}{125} & \frac{36}{25} \end{bmatrix}, \tag{10.63}$$

y que los valores de los rendimientos medios es $\mathbb{E}(R) = \left[\frac{54}{5} \quad \frac{38}{5} \quad \frac{19}{2}\right]^t$*. Determine las ponderaciones en cada una de las categorías de inversión para que el riesgo sea mínimo.*

Solución.

En este caso, se establece primero la matriz K, para ello

$$\mathbb{E}(\vec{r})\mathbf{1}^t = \begin{bmatrix} \frac{54}{5} \\ \frac{38}{5} \\ \frac{19}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{54}{5} & \frac{54}{5} & \frac{54}{5} \\ \frac{38}{5} & \frac{38}{5} & \frac{38}{5} \\ \frac{19}{2} & \frac{19}{2} & \frac{19}{2} \end{bmatrix}$$

$$\mathbf{1}\mathbb{E}(\vec{r})^t = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{54}{5} & \frac{38}{5} & \frac{19}{2} \end{bmatrix} = \begin{bmatrix} \frac{54}{5} & \frac{38}{5} & \frac{19}{2} \\ \frac{54}{5} & \frac{38}{5} & \frac{19}{2} \\ \frac{54}{5} & \frac{38}{5} & \frac{19}{2} \end{bmatrix}.$$

Por lo tanto,

$$K = \begin{bmatrix} \frac{54}{5} & \frac{54}{5} & \frac{54}{5} \\ \frac{38}{5} & \frac{38}{5} & \frac{38}{5} \\ \frac{19}{2} & \frac{19}{2} & \frac{19}{2} \end{bmatrix} - \begin{bmatrix} \frac{54}{5} & \frac{38}{5} & \frac{19}{2} \\ \frac{54}{5} & \frac{38}{5} & \frac{19}{2} \\ \frac{54}{5} & \frac{38}{5} & \frac{19}{2} \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 0 & 32 & 13 \\ -32 & 0 & -19 \\ -13 & 19 & 0 \end{bmatrix}.$$

Nótese que esta matriz es una matriz antisimétrica. Ahora,

$$\mathbf{\Sigma}^{-1}K = \begin{bmatrix} \frac{52}{105} & \frac{25}{126} & -\frac{38}{189} \\ \frac{25}{126} & \frac{625}{336} & -\frac{125}{252} \\ -\frac{38}{189} & -\frac{125}{252} & \frac{55}{63} \end{bmatrix} \begin{bmatrix} 0 & \frac{16}{5} & \frac{13}{10} \\ -\frac{16}{5} & 0 & -\frac{19}{10} \\ -\frac{13}{10} & \frac{19}{10} & 0 \end{bmatrix} = \begin{bmatrix} -\frac{353}{945} & \frac{5683}{4725} & \frac{1681}{6300} \\ -\frac{2675}{504} & -\frac{155}{504} & -\frac{6605}{2016} \\ \frac{19}{42} & \frac{1919}{1890} & \frac{5149}{7560} \end{bmatrix}.$$

Luego, la ecuación característica es

$$-\lambda\left(\lambda^2 + \frac{754469}{81648}\right) = 0.$$

Para evitar trabajar con complejos, se puede elevar la matriz $\mathbf{\Sigma}^{-1}K$ al cuadrado y calcular los valores propios y vectores propios respectivos, en este caso

$$\left(\mathbf{\Sigma}^{-1}K\right)^2 = \begin{bmatrix} -\frac{353}{945} & \frac{5683}{4725} & \frac{1681}{6300} \\ -\frac{2675}{504} & -\frac{155}{504} & -\frac{6605}{2016} \\ \frac{19}{42} & \frac{1919}{1890} & \frac{5149}{7560} \end{bmatrix}^2 = \begin{bmatrix} -\frac{124\,991}{20\,412} & -\frac{11\,191}{20\,412} & -\frac{105\,013}{27\,216} \\ \frac{58\,045}{27\,216} & -\frac{87\,233}{9072} & -\frac{71\,851}{27\,216} \\ -\frac{8930}{1701} & \frac{4712}{5103} & -\frac{223\,877}{81\,648} \end{bmatrix},$$

luego, la ecuación característica es

$$-\lambda\left[\lambda + \frac{754469}{81648}\right]^2 = 0,$$

y al calcular el vector propio correspondiente a $\lambda_1 = 0$, se tiene

$$\vec{v}_1^{\,t} = \begin{bmatrix} 19 & 13 & -32 \end{bmatrix},$$

y los vectores propios correspondientes a $\lambda_2 = -\dfrac{754\,469}{81\,648}$ son

$$\vec{v}_2 = \begin{bmatrix} 124 \\ 705 \\ 0 \end{bmatrix} \qquad \text{y} \qquad \vec{v}_3 = \begin{bmatrix} 5527 \\ 0 \\ 4465 \end{bmatrix}.$$

Luego, $\vec{w}$ debe ser una combinación lineal de los vectores propios asociados a este valor propio no nulo, por lo tanto,

$$\begin{aligned} W &= \begin{bmatrix} 124 & 5527 \\ 705 & 0 \\ 0 & 4465 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 124 & 705 & 0 \\ 5527 & 0 & 4465 \end{bmatrix} \\ &= 235 \begin{bmatrix} 0 & -16581 & 2356 \\ 16581 & 0 & 13395 \\ -2356 & -13395 & 0 \end{bmatrix}. \end{aligned}$$

Al realizar el producto dado en (10.59), se obtiene

$$\vec{w} = \frac{2}{754\,469} \begin{bmatrix} 0 & -16581 & 2356 \\ 16581 & 0 & 13395 \\ -2356 & -13395 & 0 \end{bmatrix} \begin{bmatrix} 9 \\ -7 \\ \frac{5}{2} \end{bmatrix} = \frac{1}{754\,469} \begin{bmatrix} 243\,914 \\ 365\,433 \\ 145\,122 \end{bmatrix},$$

es decir, la cartera óptima en que el inversor debe invertir está formada aproximadamente por un 32.33 % en acciones, un 48.44 % en bonos y un 19.23 % en activos inmobiliarios. Con estas ponderaciones se puede ver cómo existen claras ventajas en la diversificación de la cartera, es decir, combinando los activos disponibles, se logra disminuir parte del riesgo (varianza). El riesgo de la cartera óptima es

$$\begin{aligned} \vec{w}^t \boldsymbol{\Sigma} \vec{w} &= \begin{bmatrix} \frac{243\,914}{754\,469} & \frac{365\,433}{754\,469} & \frac{145\,122}{754\,469} \end{bmatrix} \begin{bmatrix} \frac{9}{4} & -\frac{3}{25} & \frac{9}{20} \\ -\frac{3}{25} & \frac{16}{25} & \frac{42}{125} \\ \frac{9}{20} & \frac{42}{125} & \frac{36}{25} \end{bmatrix} \begin{bmatrix} \frac{243\,914}{754\,469} \\ \frac{365\,433}{754\,469} \\ \frac{145\,122}{754\,469} \end{bmatrix} \\ &= \frac{49\,001\,031}{94\,308\,625} = 0.51958. \end{aligned}$$

Nótese que la varianza de la cartera óptima $\big(0.51958\big)$ es inferior a las otras tres varianzas de los intrumentos de inversión considerados por sí solos (ver diagonal de la matriz de varianza-covarianzas).

Este problema puede solucionarse mediante métodos numéricos, empleando un modelo de programación no lineal (NLP), puesto que la función objetivo fue expresada empleando una forma cuadrática.

Apéndice A

Métodos iterativos para estimar valores propios y vectores propios

En el Capítulo 2 se estudio un procedimiento para encontrar los valores propios de una matriz $A = [a_{ij}]$, que era resolver la ecuación característica asociada. En muchos problemas prácticos, obtener las raíces correspondientes del polinomio $p_A(\lambda)$ no es sencillo. Es más, en algunos problemas estadísticos solo se necesita el valor propio con el valor absoluto más grande. En este apéndice se trataran algunos métodos para calcular valores aproximados de los valores propios de una matriz.

A.1 Valor propio dominante y vector propio dominante

Definición A.1 Valor propio dominante y vector propio dominante

La matriz A de tamaño $n \times n$ tiene un valor propio dominante si el valor absoluto de este es mayor que los valores absolutos de los valores propios restantes. El vector propio asociado al valor propio dominante se denomina vector propio dominante.

Ejemplo A.1 *Determine el valor propio dominante para la matriz:*

$$A = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix}$$

Solución.

Los valores propios asociados a A son -1 y 6. Por lo tanto, el valor propio dominante es 6.

Ejemplo A.2 *Determine el valor propio dominante de la matriz:*

$$A = \begin{bmatrix} 4 & 2 & -2 \\ 8 & 10 & -14 \\ 8 & 8 & -12 \end{bmatrix}.$$

Solución.

Los valores propios asociados a la matriz A son 2, -4 y 4. Por lo tanto, no hay valor propio dominante.

A.1.1 Método de la potencia

El método de potencias para aproximar valores propios es iterativo. Primero, se supone que la matriz A tiene un valor propio dominante con vectores propios dominantes. Luego, se elige un vector diferente de cero $\vec{w}_1 \in \mathbb{R}^n$. Por último, se forma la sucesión definida por

$$\begin{aligned} \vec{w}_2 &= A\vec{w}_1 \\ \vec{w}_3 &= A\vec{w}_2 = A(A\vec{w}_1) = A^2\vec{w}_1 \\ \vec{w}_4 &= A\vec{w}_3 = A(A^2\vec{w}_1) = A^3\vec{w}_1 \\ &\vdots \\ \vec{w}_{k+1} &= A\vec{w}_k = A(A^{k-1}\vec{w}_1) = A^k\vec{w}_1. \end{aligned}$$

A medida que k crece, $A^k\vec{w}_1$ se hace paralelo al vector propio dominante de A.

Teorema A.1 *Sea A una matriz real diagonalizable de tamaño $n \times n$ con valores propios reales $\lambda_1, \lambda_2, \ldots, \lambda_n$ tales que*

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|.$$

Entonces, existe un vector $\vec{w}_1$ diferente de cero de $\mathbb{R}^n$ tal que la sucesión de vectores definida por

$$A\vec{w}_1, A^2\vec{w}_1, A^3\vec{w}_1, \ldots, A^k\vec{w}_1, \ldots$$

se aproxima al vector propio dominante de A cuando k aumenta.

Demostración.

Como A es diagonalizable, entonces existe una base de $\mathbb{R}^n$ formada por los n vectores propios $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ asociados a los valores propios λ_i, $i = 1, 2, \ldots, n$, respectivamente.

Sea $\vec{w}_1$ cualquier vector distinto de cero de $\mathbb{R}^n$, de forma que

$$\vec{w}_1 = \sum_{i=1}^{n} c_i \vec{v}_i, \qquad \text{con} \qquad c_i \in \mathbb{R}. \tag{A.1}$$

Definamos el siguiente proceso iterativo:

$$\vec{w}_{k+1} = A\ \vec{w}_k, \qquad k = 1, 2, \ldots$$

Nótese que $\vec{w}_{k+1} = A^k \vec{w}_1,\ k = 1, 2, \ldots$ Por lo tanto, se tiene que

$$\begin{aligned}
\vec{w}_2 = A\ \vec{w}_1 &= A\left[\sum_{i=1}^{n} c_i \vec{v}_i\right] = \sum_{i=1}^{n} c_i \lambda_i \vec{v}_i \\
&= \lambda_1 \left[c_1 \vec{v}_1 + \sum_{i=2}^{n} c_i \frac{\lambda_i}{\lambda_1} \vec{v}_i\right], \\
\vec{w}_3 = A\ \vec{w}_2 &= A\left[\lambda_1 \big(c_1 \vec{v}_1 + \sum_{i=2}^{n} c_i \frac{\lambda_i}{\lambda_1} \vec{v}_i\big)\right] \\
&= \lambda_1^2 \left[c_1 \vec{v}_1 + \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^2 \vec{v}_i\right],
\end{aligned}$$

en general, por recurrencia, se obtiene

$$\vec{w}_{k+1} = A\ \vec{w}_k = \lambda_1^k \left[c_1 \vec{v}_1 + \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \vec{v}_i\right].$$

Luego, con base en la hipótesis original de que λ_1 es mayor en valor absoluto que los demás valores propios, se concluye que cuando k tiende a infinito, cada una de las fracciones $\left[\frac{\lambda_i}{\lambda_1}\right]^k$ para $i > 1$ tiende a cero, pues $\left|\frac{\lambda_i}{\lambda_1}\right| < 1$. Esto implica que la aproximación

$$\vec{w}_{k+1} \cong \lambda_1^k c_1 \vec{v}_1, \qquad c_1 \neq 0,$$

mejora a medida que k es suficientemente grande. Como $\vec{v}_1$ es el vector propio asociado a λ_1, entonces es dominante; luego, cualquier múltiplo escalar de $\vec{v}_1$ también es un vector propio dominante. Así, se ha demostrado

que $A^k\vec{w}_1$ se aproxima arbitrariamente al vector propio dominante cuando k crece. ■

Como las componentes de $A^k\vec{w}_1$ pueden ser números muy grandes al aumentar k, esto conduce a un error de redondeo. Este problema se evita multiplicando $A^k\vec{w}_1$ por un escalar adecuado en cada iteración.

A continuación, se presenta un procedimiento para obtener el valor propio dominante de una matriz A.

Cálculo del valor propio dominante de A

i) Seleccione un vector arbitrario diferente de cero $\vec{w}_1$, cuya entrada más grande sea 1.

ii) Para $k = 1, 2, \ldots,$

   a) Calcule $A\vec{w}_k$.

   b) Sea μ_k el elemento de $A\vec{w}_k$ con valor absoluto más grande.

   c) Evalúe $\vec{w}_{k+1} = \frac{1}{\mu_k} A\vec{w}_k$.

iii) Para casi todas las escogencias de $\vec{w}_1$, la sucesión $\{\mu_k\}$ tiende al valor propio dominante y la sucesión $\{\vec{w}_k\}$ se aproxima al correspondiente vector propio.

Con esta metodología no hay reglas eficaces y rápidas para determinar cuántas iteraciones se deben realizar. Si se escoge el vector $\vec{w}_1$ de manera que en la expresión (A.1) el coeficiente c_1 sea cero, el método falla.

Ejemplo A.3 *Ilustrar el método de la potencia para la matriz dada en el Ejemplo 2.1, comenzando con* $\vec{w}_1^t = [1 \quad 1]$.

Solución.

Aplicando el método de la potencia de tal forma que en cada resultado no se utilicen cifras decimales para evitar redondeos, se obtiene lo siguiente:

k	1	2	3	4	5	6	7
$\vec{w}_k$	$\begin{bmatrix}1\\1\end{bmatrix}$	$\begin{bmatrix}\frac{1}{3}\\1\end{bmatrix}$	$\begin{bmatrix}\frac{7}{17}\\1\end{bmatrix}$	$\begin{bmatrix}\frac{41}{103}\\1\end{bmatrix}$	$\begin{bmatrix}\frac{247}{617}\\1\end{bmatrix}$	$\begin{bmatrix}\frac{1481}{3703}\\1\end{bmatrix}$	$\begin{bmatrix}0.4\\1\end{bmatrix}$
$A\vec{w}_k$	$\begin{bmatrix}3\\9\end{bmatrix}$	$\begin{bmatrix}\frac{7}{3}\\\frac{17}{3}\end{bmatrix}$	$\begin{bmatrix}\frac{41}{17}\\\frac{103}{17}\end{bmatrix}$	$\begin{bmatrix}\frac{247}{103}\\\frac{617}{103}\end{bmatrix}$	$\begin{bmatrix}\frac{1481}{617}\\\frac{3703}{617}\end{bmatrix}$	$\begin{bmatrix}\frac{8887}{3703}\\\frac{22217}{3703}\end{bmatrix}$	$\begin{bmatrix}2.4\\6.0\end{bmatrix}$
λ_k	-	9	$\frac{17}{3}$	$\frac{103}{17}$	$\frac{617}{103}$	$\frac{3703}{617}$	5.9997

Luego, el valor propio dominante es aproximadamente 5.9997 y el respectivo vector propio es $[0.4 \quad 1]^t$.
Las respuestas exactas son

$$\lambda_1 = 6 \qquad \text{y} \qquad \vec{v}_1^t = [0.4 \quad 1]\,.$$

Definición A.2 Cociente de Rayleigh

Sea A una matriz real diagonalizable de tamaño $n \times n$, se llama cociente de Rayleigh de A a la función real definida para cada $\vec{x} \neq \vec{0}$ como

$$r_A(\vec{x}) = \frac{\vec{x}^t A\vec{x}}{\langle \vec{x}, \vec{x}\rangle} = \frac{\vec{x}^t A\vec{x}}{\vec{x}^t\vec{x}}. \tag{A.2}$$

Aunque el cociente de Rayleight depende de la matriz, el subíndice A de r se omite si no hay confusión.

Teorema A.2 *Sea A una matriz real diagonalizable de tamaño $n\times n$. Sea $\vec{w}_1 \in \mathbb{R}^n$ cualquier vector no nulo. Considérense los cocientes de Rayleigh:*

$$r(\vec{w}_k) = \frac{\vec{w}_k^t A\vec{w}_k}{\vec{w}_k^t\vec{w}_k}, \qquad \textit{para} \qquad k = 1, 2, \ldots, m,$$

donde m es la cantidad deseada de iteraciones. El último cociente $r(\vec{w}_m)$ es una aproximación del valor propio dominante λ de A, y si se hace $r(\vec{w}_k) = \lambda + \epsilon$, de modo que ϵ es el error de $r(\vec{w}_k)$, entonces:

$$|\epsilon| \leq \sqrt{\frac{\vec{y}_k^t\ \vec{y}_k}{\vec{w}_k^t\ \vec{w}_k} - r^2(\vec{w}_k)}, \tag{A.3}$$

donde $\vec{y}_k = A\vec{w}_k$.

Demostración.

Si se reescribe el cociente de Rayleigh, se obtiene que

$$\vec{w}_k^t A\vec{w}_k = r\big(\vec{w}_k\big)\vec{w}_k^t\ \vec{w}_k, \qquad\qquad k = 1, 2, \ldots, m,$$

y dado que $\vec{y}_k = A\vec{w}_k$, se tiene que $\vec{w}_k^t\vec{y}_k = r\big(\vec{w}_k\big)\vec{w}_k^t\vec{w}_k$. Por lo tanto,

$$\begin{aligned}\big[\vec{y}_k - r\big(\vec{w}_k\big)\vec{w}_k\big]^t\big[\vec{y}_k - r\big(\vec{w}_k\big)\vec{w}_k\big] &= \vec{y}_k^t\vec{y}_k - 2r\big(\vec{w}_k\big)\vec{w}_k^t\vec{y}_k + r^2\big(\vec{w}_k\big)\vec{w}_k^t\vec{w}_k \\ &= \vec{y}_k^t\vec{y}_k - r^2\big(\vec{w}_k\big)\vec{w}_k^t\vec{w}_k \\ &= \left[\frac{\vec{y}_k^t\vec{y}_k}{\vec{w}_k^t\vec{w}_k} - r^2\big(\vec{w}_k\big)\right]\vec{w}_k^t\vec{w}_k \\ &= \epsilon^2\vec{w}_k^t\vec{w}_k. \end{aligned} \tag{A.4}$$

Como A es una matriz diagonalizable, por el Teorema 2.19 tiene n vectores propios linealmente independientes $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ correspondientes a los valores propios $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, respectivamente, y mediante el proceso de Gram-Schmidt se ortonormalizan estos vectores para obtener $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ una base ortonormal de $\mathbb{R}^n$. Entonces, $\vec{w}_k$ tiene una representación de la forma

$$\vec{w}_k = \sum_{i=1}^{n} c_i\vec{v}_i \qquad\qquad \text{con} \qquad\qquad c_i \in \mathbb{R},$$

y como los $\vec{v}_i$ son vectores unitarios ortogonales, entonces

$$\vec{w}_k^t\vec{w}_k = \sum_{i=1}^{n} c_i^2. \tag{A.5}$$

Ahora bien,

$$\vec{y}_k = A\vec{w}_k = A\left[\sum_{i=1}^{n} c_i\vec{v}_i\right] = \sum_{i=1}^{n} c_i\lambda_i\vec{v}_i.$$

Luego,

$$\vec{y}_k - r\big(\vec{w}_k\big)\vec{w}_k = \sum_{i=1}^{n} c_i\big[\lambda_i - r\big(\vec{w}_k\big)\big]\vec{v}_i.$$

Si se reemplaza en (A.4), se obtiene que

$$\epsilon^2\vec{w}_k^t\vec{w}_k = \sum_{i=1}^{n} c_i^2\big[\lambda_i - r\big(\vec{w}_k\big)\big]^2.$$

Si se sustituye cada $\left[\lambda_i - r\left(\vec{w}_k\right)\right]^2$ por el menor de estos términos y se aplica (A.5), se tiene que

$$\epsilon^2 \vec{w}_k^t \vec{w}_k \geq \left[\lambda_c - r\left(\vec{w}_k\right)\right]^2 \sum_{i=1}^{n} c_i^2 = \left[\lambda_c - r\left(\vec{w}_k\right)\right]^2 \vec{w}_k^t \vec{w}_k,$$

donde λ_c es un valor propio al cual $r\left(\vec{w}_k\right)$ está próximo. De esto se llega a (A.3), y queda demostrado el teorema. ■

A continuación, se presenta un procedimiento para obtener el valor propio dominante de una matriz A.

Método de los cocientes de Rayleigh

Sea A una matriz diagonalizable de tamaño $n \times n$ con un valor propio dominante. Sea m la cantidad deseada de iteraciones.

i) Seleccione un vector arbitrario diferente de cero $\vec{w}_0$.

ii) Para $k = 0, 1, \ldots, m-1$,

   a) Calcule $\vec{z}_k = \dfrac{\vec{w}_k}{\|\vec{w}_k\|}$.

   b) Sea $\vec{w}_{k+1} = A\vec{z}_k$.

   c) Evalúe $r\left(\vec{z}_k\right) = \vec{z}_k^t \vec{w}_{k+1}$.

iii) Los cocientes de Rayleigh $\left\{r\left(\vec{z}_k\right)\right\}$ tienden al valor propio dominante y la sucesión $\{\vec{z}_k\}$ se aproxima al correspondiente vector propio unitario.

Para matrices simétricas, este método es muy eficiente y requiere menos iteraciones para lograr la misma exactitud.

Ejemplo A.4 *Ilustrar el método de los cocientes de Rayleigh, para la matriz dada en el Ejemplo A.3*

Solución.

Aplicando el método de los cocientes de Rayleigh se obtiene la tabla siguiente:

k	0	1	2	3	4	5
$\vec{z}_k$	$\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{bmatrix}$	$\begin{bmatrix} \frac{7}{13\sqrt{2}} \\ \frac{17}{13\sqrt{2}} \end{bmatrix}$	$\begin{bmatrix} \frac{41}{\sqrt{12290}} \\ \frac{103}{\sqrt{12290}} \end{bmatrix}$	$\begin{bmatrix} \frac{247}{\sqrt{441698}} \\ \frac{617}{\sqrt{441698}} \end{bmatrix}$	$\begin{bmatrix} 0.37135 \\ 0.92849 \end{bmatrix}$
$A\vec{w}_k$	$\begin{bmatrix} \frac{3}{\sqrt{2}} \\ \frac{9}{\sqrt{2}} \end{bmatrix}$	$\begin{bmatrix} \frac{7}{\sqrt{10}} \\ \frac{17}{\sqrt{10}} \end{bmatrix}$	$\begin{bmatrix} \frac{41}{13\sqrt{2}} \\ \frac{103}{13\sqrt{2}} \end{bmatrix}$	$\begin{bmatrix} \frac{247}{\sqrt{12290}} \\ \frac{617}{\sqrt{12290}} \end{bmatrix}$	$\begin{bmatrix} \frac{1481}{\sqrt{441698}} \\ \frac{3703}{\sqrt{441698}} \end{bmatrix}$	$\begin{bmatrix} 2.2283 \\ 5.5707 \end{bmatrix}$
λ_k	-	$\frac{29}{5}$	$\frac{1019}{169}$	$\frac{36839}{6145}$	$\frac{1325279}{220849}$	5.9997

Así, el valor propio dominante es aproximadamente $\lambda = 6$ y el respectivo vector propio unitario es $\begin{bmatrix} 0.37135 \\ 0.92849 \end{bmatrix}$.

Ejercicios A.1 Determine los vectores propios dominantes con los métodos descritos en este apéndice para las siguientes matrices:

$$a. \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}. \qquad b. \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}. \qquad c. \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

$$d. \begin{bmatrix} 1 & 3 & -2 \\ 3 & 1 & 2 \\ -2 & 2 & 1 \end{bmatrix}. \qquad e. \begin{bmatrix} 3 & -1 & 4 \\ -1 & 1 & -5 \\ 4 & -5 & 2 \end{bmatrix}.$$

Apéndice B

Números complejos

B.1 Álgebra de los números complejos

Como el concepto de número complejo se utiliza con mucha frecuencia en este material y como algunos lectores quizás tengan solo un conocimiento superficial de ellos, este apéndice contiene un breve repaso de sus propiedades algebraicas más importantes.

Definición B.1 Número complejo

Un número complejo es una expresión de la forma:

$$z = a + bi,$$

donde a y b son números reales: a se llama la parte real de z y se denota por $Re(z)$ y b es llamado la parte imaginaria de z y lo denotamos por $Im(z)$. El símbolo i se llama unidad imaginaria y satisface la propiedad de que $i^2 = -1$.

Definición B.2 Igualdad de números complejos

Dos números complejos $a + bi$ y $c + di$ se definen como iguales si y solo si las partes real e imaginaria de uno son respectivamente iguales a las partes real e imaginaria del otro, esto es, si y solo si $a = c$ y $b = d$.

Definición B.3 Formas especiales de los números complejos

Dado un número complejo de la forma $z = a + bi$, si $b = 0$, se llama número real; por otra parte, si $a = 0$, se denomina número imaginario puro.

B.1.1 Operaciones fundamentales

Definición B.4 Suma y diferencia

La suma y diferencia de los números complejos $a + bi$ y $c + di$ son definidas sumando o restando sus partes reales y sus partes imaginarias, como sigue:

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$
$$(a + bi) - (c + di) = (a - c) + (b - d)i.$$

Definición B.5 Multiplicación

El producto de los números complejos $a + bi$ y $c + di$ se define como sigue:

$$(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i. \tag{B.1}$$

Teorema B.1 *Un número complejo es igual a cero si y solo si sus partes real e imaginaria valen cero.*

Demostración.

Queda como ejercicio para el lector. ■

Definición B.6 Inverso

El recíproco o inverso del número complejo $c+di$ se define como sigue

$$\big(c + di\big)^{-1} = \frac{c}{c^2 + d^2} + \frac{-d}{c^2 + d^2}i = \frac{c - di}{c^2 + d^2}.$$

Puesto que c^2 y d^2 son no negativos, $c^2 + d^2 = 0$ si y solo si $c = d = 0$. Por lo tanto, el único número complejo $c + di$ que no tiene recíproco es el cero.

Esta definición del recíproco nos lleva a la siguiente definición.

Definición B.7 División

Si $w = a + bi$, $z = c + di$ y $z \neq 0$, se puede definir su cociente como sigue:

$$\frac{w}{z} = w.z^{-1} = \left(\frac{ac + bd}{c^2 + d^2}\right) + \left(\frac{bc - ad}{c^2 + d^2}\right)i, \tag{B.2}$$

el cual resulta ser un número complejo.

Teorema B.2 *Los números complejos satisfacen las siguientes propiedades:*

- *Para la adición:*

  A1. Si $z_1 = a + bi$ y $z_2 = c + di$ pertenecen a $\mathbb{C}$, su suma $z_1 + z_2$ pertenece a $\mathbb{C}$. Esto también se expresa diciendo que $\mathbb{C}$ es cerrado bajo la adición.

  A2. Ley Conmutativa: $z_1 + z_2 = z_2 + z_1$.

  A3. Ley Asociativa: $(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$.

  A4. Elemento identidad: Existe un elemento $0 = 0 + 0i$ en $\mathbb{C}$ tal que si $z \in \mathbb{C}$, $z + 0 = z$.

  A5. Si $z \in \mathbb{C}$, existe un elemento único $-z$ en $\mathbb{C}$, llamado el negativo de z, tal que $z + (-z) = 0$.

- *Para la multiplicación:*

  M1. Si $z_1 = a + bi$ y $z_2 = c + di$ pertenecen a $\mathbb{C}$, su producto $z_1.z_2$ pertenece a $\mathbb{C}$. Esto también se expresa diciendo que $\mathbb{C}$ es cerrado bajo la multiplicación.

  M2. Ley Conmutativa: $z_1.z_2 = z_2.z_1$.

  M3. Ley Asociativa: $(z_1.z_2).z_3 = z_1.(z_2.z_3)$.

  M4. Elemento identidad: Existe un elemento $1 = 1 + 0i$ en $\mathbb{C}$ tal que $1.z = z$ para todo $z \in \mathbb{C}$.

  M5. Si $z \neq 0$, existe un elemento único z^{-1} tal que $z.(z^{-1}) = 1$.

- *Ley Distributiva: Esta última regla entrelaza la adición y la multiplicación.*

  Si z_1, z_2 y z_3 pertenecen a $\mathbb{C}$, entonces

$$z_1.(z_2 + z_3) = (z_1.z_2) + (z_1.z_3).$$

Demostración.
Queda como ejercicio para el lector. ■

Definición B.8 Número complejo conjugado
El conjugado de z es el número $\overline{z} = \overline{a + bi} = a - bi$.

Teorema B.3 *La suma, diferencia y producto de números complejos con sus conjugados son, respectivamente, un número real, un número imaginario puro y un número real no negativo.*

Demostración.

Si $z = a + bi$, entonces $\overline{z} = a - bi$ y por lo tanto,

$$z + \overline{z} = 2a, \qquad z - \overline{z} = 2bi \qquad \text{y} \qquad z.\overline{z} = a^2 + b^2.$$

■

Teorema B.4 *Si un número complejo es igual a su conjugado es un número real, pero si es igual al negativo de su conjugado es un número imaginario puro.*

Demostración.

Si $z = a + bi$, entonces $\overline{z} = a - bi$, luego si $z = \overline{z}$, por la definición de igualdad se tiene que $b = -b$, así que $b = 0$, y por lo tanto $z = a$. Por otra parte, si $z = -\overline{z}$, por la definición de igualdad se tiene que $a = -a$, de modo que $a = 0$ y $z = bi$. ■

Una aplicación importante del conjugado de un número complejo está en el cálculo de un cociente, la regla *"multiplíquense numerador y denominador por el conjugado del denominador"* es más fácil de recordar que la fórmula (B.2). En otras palabras, en el proceso de división

$$z^{-1} = \frac{1}{z} = \frac{\overline{z}}{z.\overline{z}}$$

y en forma análoga

$$\frac{w}{z} = w.z^{-1} = \frac{w.\overline{z}}{z.\overline{z}}, \qquad z \neq 0.$$

Teorema B.5 Propiedades de los números complejos conjugados

Si w y z son números complejos, entonces

(a) $\overline{\overline{w}} = w$. (b) $\overline{w \pm z} = \overline{w} \pm \overline{z}$.

(c) $\overline{w.z} = \overline{w}.\overline{z}$. (d) $(\overline{w/z}) = \overline{w}/\overline{z}$, *si* $z \neq 0$.

Demostración.

(c) Si $w = a + bi$ y $z = c + di$, entonces por (B.1)

$$\begin{aligned} \overline{w.z} &= \overline{(a + bi).(c + di)} = \overline{(ac - bd) + (ad + bc)i} \\ &= (ac - bd) - (ad + bc)i. \end{aligned}$$

Mientras que

$$\begin{aligned}\overline{w}.\overline{z} &= \overline{(a+bi)}.\overline{(c+di)} = (a-bi).(c-di)\\ &= (ac-bd)-(ad+bc)i.\end{aligned}$$

De modo que

$$\overline{w.z} = \overline{w}.\overline{z}.$$

Procedimientos semejantes se aplican en los otros casos. ■

Definición B.9 Módulo

El módulo o valor absoluto del número complejo $z = a + bi$, representado por $|z|$, es la distancia desde el punto (a, b) al origen, es decir:

$$|z| = \sqrt{z.\overline{z}} = \sqrt{a^2+b^2}.$$

Teorema B.6 Propiedades del módulo

Si w y z son números complejos, entonces

(a) $|\overline{w}| = |w|$.

(b) $|w+z| \leq |w| + |z|$.

(c) $|w.z| = |w|.|z|$.

(d) $\left|\dfrac{w}{z}\right| = \dfrac{|w|}{|z|}$, *si* $z \neq 0$.

Demostración.

Sean $w = a + bi$ y $z = c + di$, entonces

(a) $|\overline{w}| = \sqrt{a+(-b)^2} = |w|$.

(b) Para probar esta, observemos que

$$\begin{aligned}(w+z).\overline{(w+z)} &= (w+z).(\overline{w}+\overline{z})\\ &= w.\overline{w} + z.\overline{z} + \overline{w}.z + w.\overline{z},\end{aligned}$$

o bien

$$|w+z|^2 = |w|^2 + |z|^2 + \left(\overline{w}.z + \overline{\overline{w}.z}\right).$$

Pero como

$$\overline{w}.z + \overline{\overline{w}.z} = 2Re\,(\overline{w}.z) \leq |\overline{w}.z| = 2|w|.|z|,$$

se tiene que

$$|w+z|^2 \leq |w|^2 + |z|^2 + 2|w|.|z| = (|w| + |z|)^2.$$

Tomando raíz cuadrada en ambos miembros, se llega al resultado deseado:

$$|w+z| \leq |w| + |z|.$$

(c) La demostración consiste en un cálculo directo:

$$\begin{aligned} |w.z| &= |(a+bi).(c+di)| = |(ac-bd)+(ad+bc)i| \\ &= \sqrt{(ac-bd)^2+(ad+bc)^2} = \sqrt{(a^2+b^2).(c^2+d^2)} \\ &= |w|.|z|. \end{aligned}$$

(d) Se deja la prueba para el lector. ■

Definición B.10 Argumento

El argumento o amplitud del número complejo $z = a + bi$ es el ángulo formado por el segmento que va del origen al punto que representa un número complejo y el eje real positivo, y está dado por la expresión:

$$\begin{aligned} \theta = \arg(z) &= \underbrace{\arctan\left(\frac{b}{a}\right)}_{} + 2n\pi \qquad\qquad n = 0, \pm 1, \pm 2, \ldots \\ &= \quad Arg(z) \quad + 2n\pi, \end{aligned} \tag{B.3}$$

donde $Arg(z)$ denota el valor principal de $\arg(z)$ *y se define como el único valor de* $\arg(z)$ *tal que* $-\pi \leq \arg(z) < \pi$.

Teorema B.7 Propiedades del argumento

Si w y z son números complejos, entonces

(a) $\arg(w.z) = \arg(w) + \arg(z)$.

(b) $\arg\left(\frac{w}{z}\right) = \arg(w) - \arg(z)$, *si* $z \neq 0$.

Demostración.

Queda como ejercicio para el lector. ■

B.1.2 Representación polar

Sean r y θ coordenadas polares del punto (a, b) que corresponden a un número complejo no nulo $z = a + bi$. Vea la representación en la Figura B.1:

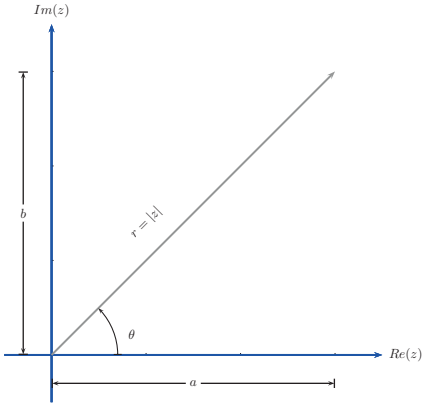

Figura B.1: Representación polar.

Por tanto,

$$a = r\cos\theta = |z|\cos(\arg z) \qquad \text{y} \qquad b = r\,\mathrm{sen}\,\theta = |z|\,\mathrm{sen}\,(\arg z)\,.$$

En consecuencia, z puede ser expresado en forma polar como

$$z = r\big(\cos\theta + i\,\mathrm{sen}\,\theta\big) = |z|\big[\cos\big(\arg z\big) + i\,\mathrm{sen}\big(\arg z\big)\big].$$

Esta representación polar de z es de gran utilidad para obtener potencias y raíces de números complejos.

Definición B.11 Fórmula de Euler

Sea θ un número real. Se define el símbolo $e^{i\theta}$, como sigue:

$$e^{i\theta} = \cos\theta + i\,\mathrm{sen}\,\theta, \tag{B.4}$$

esta ecuación se conoce como la fórmula de Euler.

Teorema B.8 Teorema de De Moivre

Si $z = r\big(\cos\theta + i\,\mathrm{sen}\,\theta\big)$, entonces

$$z^n = r^n\big(\cos n\theta + i\,\mathrm{sen}\,n\theta\big) = |z|^n\big[\cos\big(n\arg z\big) + i\,\mathrm{sen}\big(n\arg z\big)\big], \tag{B.5}$$

donde n es cualquier número entero.

Demostración.

Queda como ejercicio para el lector. ■

El teorema de De Moivre también puede utilizarse para encontrar las raíces m-ésimas de un número complejo. Si se hace $n = \frac{1}{m}$, entonces

$$z^{\frac{1}{m}} = \left[r\left(\cos\theta + i\,\mathrm{sen}\,\theta\right)\right]^{\frac{1}{m}} = r^{\frac{1}{m}}\left[\cos\left(\frac{1}{m}\theta\right) + i\,\mathrm{sen}\left(\frac{1}{m}\theta\right)\right].$$

Además, si se tienen en cuenta las siguientes identidades trigonométricas

$$\cos\big(\theta + 2k\pi\big) = \cos\theta \qquad \text{y} \qquad \mathrm{sen}\big(\theta + 2k\pi\big) = \mathrm{sen}\,\theta,$$

en donde k es un entero, se tiene que

$$\begin{aligned} z^{\frac{1}{m}} &= \left\{r\left(\cos\left(\theta + 2k\pi\right) + i\,\mathrm{sen}\left(\theta + 2k\pi\right)\right)\right\}^{\frac{1}{m}} \\ &= r^{\frac{1}{m}}\left[\cos\left(\frac{\theta + 2k\pi}{m}\right) + i\,\mathrm{sen}\left(\frac{\theta + 2k\pi}{m}\right)\right]. \end{aligned}$$

Las raíces m-ésimas se obtienen asignando a k los m valores consecutivos enteros $0, 1, 2, \ldots, m-1$.

Observación.
Si $m = 2$, se tiene que

$$\sqrt{z} = \sqrt{a \pm bi} = |z|^{\frac{1}{2}}\left[\cos\left(\frac{\theta}{2} + k\pi\right) \pm i\,\mathrm{sen}\left(\frac{\theta}{2} + k\pi\right)\right], \quad k = 0, 1.$$

Luego,

$$\sqrt{z} = \begin{cases} |z|^{\frac{1}{2}}\left[\cos\left(\frac{\theta}{2}\right) \pm i\,\mathrm{sen}\left(\frac{\theta}{2}\right)\right] & \text{si} \quad k = 0, \\ -|z|^{\frac{1}{2}}\left[\cos\left(\frac{\theta}{2}\right) \pm i\,\mathrm{sen}\left(\frac{\theta}{2}\right)\right] & \text{si} \quad k = 1. \end{cases}$$

Si se usan las siguientes identidades trigonométricas

$$\cos\left(\frac{\theta}{2}\right) = \sqrt{\frac{1 + \cos\theta}{2}} \qquad \text{y} \qquad \mathrm{sen}\left(\frac{\theta}{2}\right) = \sqrt{\frac{1 - \cos\theta}{2}},$$

se llega a

$$\sqrt{z} = (-1)^k\,\frac{\sqrt{2}}{2}\left[\sqrt{a + |z|} + \mathrm{sgn}\,(b)\,\sqrt{a - |z|}\right], \qquad k = 0, 1, \qquad \text{(B.6)}$$

donde $\mathrm{sgn}(\cdot)$ denota la función signo.

Ejemplo B.1 *Si $z = a + bi$, demuestre que*

$$\sqrt{z} \pm \sqrt{\overline{z}} = \pm\sqrt{2\,(a \pm |z|)}.$$

Solución.

Si $z = a + bi$, entonces $\overline{z} = a - bi$. Por lo tanto,

$$\left(\sqrt{z} \pm \sqrt{\overline{z}}\right)^2 = z + \overline{z} \pm 2\sqrt{z.\overline{z}}.$$

Por el teorema B.3, se tiene

$$\left(\sqrt{z} \pm \sqrt{\overline{z}}\right)^2 = 2a \pm 2\sqrt{a^2 + b^2}.$$

Al tomar raíz cuadrada a ambos lados, se obtiene

$$\sqrt{z} \pm \sqrt{\overline{z}} = \pm\sqrt{2\,(a \pm |z|)},$$

y la prueba queda completa.

Bibliografía

Anton, H. (1996), *Introducción al Álgebra Lineal*, Editorial Limusa S.A. Grupo Noriega Editores, México.

Apostol, T. M. (1985), *Calculus*, Vol. 2, 2^a edn, Editorial Reverté S.A., Barcelona.

Asmar, A. J. (1995), *Tópicos en Teoría de Matrices*, Universidad Nacional de Colombia, sede Medellín, Colombia.

Barbolla, S. & Sanz, P. (1998), *Álgebra Lineal y Teoría de Matrices*, Prentice Hall, Madrid.

Bernstein, D. S. (2009), *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd edn, Princeton University Press, Estados Unidos.

Bretscher, O. (1997), *Linear Algebra with Applications*, Prentice Hall, New Jersey.

Bru, R., Climent, J.-J., Mas, J. & Urbano, A. (2001), *Álgebra Lineal*, Alfaomega S.A., México.

Cayley, A. (1858), 'A memoir on the theory of matrices', *Philosophical Transactions of the Royal Society of London* **148**, 17-37.

Ding, Z. & Engle, R. (2001), 'Large scale conditional covariance matrix modeling estimation and testing', *Academia Economic Papers* **29**(2), 157-184.

Filippov, A. F. (1971), 'A short proof of the theorem on reduction of a matrix to jordan form', *Moscow University Mathematics Bulletin* **26**, 70-71.

Fraleigh, B. (1989), *Álgebra Lineal*, Addison-Wesley Iberoamericana S.A., Estados Unidos.

Frazer, R. A., Duncan, W. J. & Collar, A. R. (1965), *Elementary Matrices and Some Applications to Dynamics and Differential Equations*, The Syndics of the Cambridge University Press, Estados Unidos.

Golubitsky, M. & Dellnitz, M. (2001), *Álgebra Lineal y Ecuaciones Diferenciales con Matlab*, International Thomson Editores S.A., México.

Graybill, F. A. (1983), *Matrices with applications in statistics*, Cole statistics / probability series, 2nd edn, Wadsworth & Brooks, Estados Unidos.

Grossman, S. I. (1996), *Álgebra Lineal*, 5^a edn, McGraw-Hill, México.

Grossman, S. I. (1998), *Aplicaciones de Álgebra Lineal*, Grupo Editorial Iberoamérica S.A., México.

Herstein, I.Ñ. & Winter, D. (1989), *Álgebra Lineal y Teoría de Matrices*, Grupo Editorial Iberoamérica S.A., México.

Hoaglin, D. C. & Welsch, R. E. (1978), 'The hat matrix in regression and anova', *The American Statistician* **32**(1), 17-22.

Horn, R. & Johnson, C. (1985), *Matrix Analysis*, Cambridge University Press, Estados Unidos.

Kemeny, J. G. & Snell, J. L. (1976), *Finite Markov Chains*, Springer-Verlag, Nueva York.

Kolman, B. (1997), *Álgebra Lineal con Aplicaciones y matlab*, Prentice Hall, México.

Lang, S. (2004), *Linear algebra*, 3rd edn, Springer, Estados Unidos.

Larson, E. (2000), *Introducción al Álgebra Lineal*, Limusa Noriega Editores, México.

Lay, D. C. (1994), *Linear Algebra and its Applications*, Addison-Wesley Publishing Company, Estados Unidos.

Markowitz, H. M. (1952), 'Portfolio selection', *The Journal of Finance* **7**(1), 77-91.

Markowitz, H. M. (1959), *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, New York.

Mendel, G. J. (1865), 'Versuche über pflanzen-hybriden', *Verhandlungen des naturforschenden Vereines, Abhandlungen, Brünn* **IV**(1), 3-47. An English translation, "Experiments in Plant Hybridisation", was published in the *Journal of the Royal Horticultural Society*, London, 1901, **26**, 1-32.

Merton, R. C. (1972), 'An analytic derivation of the efficient portfolio frontier', *The Journal of Financial and Quantitative Analysis* **7**(4), 1851-1872.

Murdoch, D. C. (1970), *Linear Algebra*, John Wiley & Sons, Nueva York.

Nakos, G. & Joyner, D. (1999), *Álgebra Lineal con Aplicaciones*, International Thomson Editores S.A., México.

Noble, B. & Daniel, J. W. (1989), *Álgebra Lineal Aplicada*, Prentice-Hall Hispanoamericana S.A., México.

Paige, L. & Swift, D. (1961), *Elements of Linear Algebra*, Blaisdell Publishing Company, Massachusetts.

Penrose, R. (1955*a*), 'A generalized inverse for matrices', *Proceedings of the Cambridge Philosophical Society* **51**, 406-413.

Penrose, R. (1955*b*), 'On best approximate solutions of linear matrix equations', *Proceedings of the Cambridge Philosophical Society* **52**, 17-19.

Pringle, R. M. & Rayner, A. A. (1971), *Generalized Inverse Matrices with Applications to Statistics*, Charles Griffin & Company Limited, Londres.

Rao, C. R. & Rao, M. B. (1998), *Matrix Algebra and its Applications to Statistics and Econometrics*, World Scientific, Estados Unidos.

Rorres, C. & Anton, H. (1979), *Aplicaciones de Álgebra Lineal*, Limusa, México.

Samelson, H. (1974), *An Introduction to Linear Algebra*, John Wiley & Sons, Nueva York.

Schott, J. R. (1997), *Matrix Analysis for Statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Estados Unidos.

Searle, S. (1971), *Linear Models*, John Wiley & Sons, Nueva York.

Searle, S. (1982), *Matrix Algebra Useful for Statistics*, John Wiley & Sons, Estados Unidos.

Strang, G. (1986), *Álgebra Lineal y sus Aplicaciones*, Addison Wesley Iberoamericana, México.

Sylvester, J. J. (1973), *The Collected Mathematical Papers*, Vol. I–IV, Chelsea Publishing Company, New York.

Weintraub, S. H. (2009), *Jordan Canonical Form: Theory and Practice*, Synthesis Lectures on Mathematics and Statistics, Morgan & Claypool Publishers series, Estados Unidos.

Zhang, F. (2005), *The Schur Complement and Its Applications (Numerical Methods and Algorithms)*, Springer Science + Business Media, Inc., Estados Unidos.

Índice alfabético

Adjunta de una matriz, 16
Amplitud de un número complejo, 464
Ángulo(s)
  de rotación, 280
  eulerianos, 294
Argumento
  de un número complejo, 464
  principal, 464
Autovalor, 70
Autovector, 70

Base, 47
  cambio de, 49
Bloque
  de Jordan, 154

Cadena de Markov, 411
Cambio de base, 49
Clasificación
  de formas cuadráticas, 263
  de formas hermíticas, 310
Cociente
  de Rayleigh, 455
Combinación lineal, 46
Complemento
  de Schur, 36
  Ortogonal, 53

Delta de Kronecker, 21
Descomposición
  de Cholesky, 183
  de Schur, 136
  de Sylvester, 84, 209
  en valores singulares, 189, 223
  espectral
    para hermitianas, 217
    para simétricas, 112
  LS, 128
  polar, 224, 225
    a derecha, 192
    a izquierda, 192
Desigualdad
  de Fischer, 270
  triangular, 316
Determinante
  de una matriz 2×2, 12
  de una matriz 3×3, 12
  de una matriz $n \times n$, 14
Diagonalización
  de una forma cuadrática, 244
  de una forma hermítica, 305
  por completación de cuadrados, 244
  por transformación
    ortogonal, 255
    unitaria, 308

Ecuación
  característica, 73
  cuadrática, 271

Ejes principales, 256
Espacio
de las columnas de una matriz, 54
de los renglones de una matriz, 54
generado, 47
nulo, 54
propio, 75
generalizado, 116
vectorial, 44
real, 65

Factorización
de una matriz, 127
LDU, 132
LU, 135
QR, 143
Forma(s)
bilineal, 231
alternada, 235
rango, 234
simétrica, 235
canónica de Jordan, 159
cuadrática(s), 238
clasificación, 263
equivalentes, 241
interpretación geométrica, 271
hermíticas
clasificación, 310
polar, 241
sesquilineal, 299
Fórmula
de Euler, 465
de Schur, 36
o expansión de Laplace, 14

g-inversa, 364

Índice
de condición, 329
Inversa
generalizada, 364
de Penrose, 391
Isomorfismo, 61

Lema
de Banach, 321
Ley de la inercia
de Sylvester, 259
Lugares
geométricos, 284, 296

Matrices
complejas, 61
semejantes, 208
congruentes, 88
hermitianas, 209
ortogonalmente, 105
particionadas, 25–42
semejantes, 90
ortogonalmente, 105
Matriz
adjunta, 16
antihermitiana, 203
antisimétrica, 21
propiedades, 21
compleja, 61
conjugada, 61
de cofactores, 13
de Jordan, 155
de permutación, 23
de reflexión, 23
de rotación, 279, 294
de transformación, 60
de transición, 412
de una forma cuadrática, 238
definida
negativa, 264
positiva, 263
determinante de una, 12

diagonal, 19
propiedades, 19
diagonalizable, 93
escalar, 18
escalonada, 8
estocástica, 409
hat, 431
propiedades, 432
hermitiana, 198
Householder, 337
idempotente, 333
identidad, 5
indefinida, 264
inversa, 10
generalizada, 364
menor de una, 13
nilpotente, 24
normal, 207
ortogonal
impropia, 23
propia, 23
propiedades, 22
regular, 411
simétrica, 17
propiedades, 18
transpuesta
conjugada, 63
transpuesta de una, 6
triangular, 17
propiedades, 17
triangularizable, 136
unitaria, 205
Método
de Cayley, 163
de eliminación de Gauss, 57
de Gauss-Jordan
Cálculo de inversa, 10
de la potencia, 452
de mínimos cuadrados, 429
de Penrose, 391
de reducción de Lagrange, 249
Módulo
de un complejo, 463
Multicolinealidad, 433
Multiplicidad
algebraica, 73
geométrica, 76

Norma
de un vector, 51
propiedades, 51
de una matriz, 315
tipos, 317–320
Núcleo, 54
Número
complejo, 459
conjugado, 461
módulo, 463
de condición, 324

Operador
Vec, 353
propiedades, 354

Polinomio
característico, 73
de matriz, 148
mínimo, 148, 150
Probabilidad
de transición, 411
Proceso
de Markov, 411
Producto
Hadamard, 355
propiedades, 355
Kronecker, 348
propiedades, 349
punto
en $\mathbb{C}^n$, 65
en $\mathbb{R}^n$, 50

Radio
espectral, 320
Raíz cuadrada
de un número complejo, 466
de una matriz, 162–181
diagonal, 164
triangular, 167
Rango
de una forma bilineal, 234
de una matriz, 55
propiedades, 55
Regla
de Cramer, 58
de Sarrus, 12
Rotación de ejes
en $\mathbb{R}^2$, 277
en $\mathbb{R}^3$, 290

Seudoinversa, 364
Sistema de ecuaciones
de mal comportamiento, 324
lineales, 56
consistente, 57
homogéneo, 57
inconsistente, 57
Submatriz
angular, 29
principal, 28
Suma
de subespacios, 45
directa, 46
Superficie
cuádrica, 284

Teorema
de Cayley-Hamilton, 149
de Cochran, 345
de De Moivre, 465
de Euler, 261
de los ejes principales, 256
de Schur, 210
espectral
para matrices simétricas, 113
Transformación
biyectiva, 61
lineal, 59
representación matricial, 59
sobre, 61
uno a uno, 60
ortogonal, 277
Transpuesta
de una matriz, 6
conjugada, 63
particionada, 30
propiedades, 6
Traza
de una matriz, 9

Valor
característico, 70
propio
dominante, 451
singular, 186
Vector(es)
característico, 70
complejo, 66
linealmente
dependientes, 46
independientes, 47
probabilístico, 412
propio
dominante, 451
generalizado, 115

El álgebra de matrices es, en la actualidad, un elemento fundamental de los conocimientos matemáticos para ingenieros y científicos; además, la comprensión de sus métodos fundamentales, es de gran utilidad para sociólogos, economistas, estudiantes de pedagogía y de ciencias económicas.

Sin embargo, a pesar de las diversas aplicaciones del álgebra matricial, la mayoría de textos de álgebra lineal no las introducen y en muchos casos, no se encuentra un libro que se ajuste a los requerimientos y necesidades de ciertas disciplinas. El presente libro busca solucionar esta necesidad, capacitando al lector para que adquiera la habilidad de usar el álgebra de matrices en diferentes ámbitos de forma práctica y concisa.

colección **textos**